ABSTRACT

Title of Dissertation: Robust Optimization Model for Bus Priority under Arterial Progression

Meenakshy Vasudevan, Doctor of Philosophy, 2005

Dissertation directed by: Professor Gang-Len Chang
Department of Civil Engineering

The purpose of this study is to design a real-time robust arterial signal control system that gives priority to buses while simultaneously maximizing progression bandwidths and optimizing signal timing plans at each intersection along the arterial. The system architecture is divided into three levels. At the *progression control level* bandwidths are maximized. Existing progression strategies do not use real-time traffic data or use simple mathematical models to estimate traffic evolution. The proposed model eliminates this drawback by using real-time data to develop a neural network model for predicting traffic flows. Rather than using pre-specified values, queue clearance and minimum green times are computed as functions of the predicted queues. To eliminate uncertainty in the prediction due to the long time horizon, robust discrete optimization technique is used to determine the progression bands.

At the *intersection control level*, signal timing plans are optimized subject to bandwidth constraints to allow for uninterrupted arterial flow, and minimum green constraints for driver safety and to discharge average waiting queues.

At the *bus priority control level*, whenever a bus is detected and is a candidate for priority it is granted priority based on a performance index that is a function of bus schedule delay, automobile and bus passenger delays, and vehicle delays, subject to bandwidth and minimum green constraints. Minimum green constraints ensure that other traffic users are not unduly penalized. Bandwidth constraints allow for uninterrupted arterial flow despite a preferential treatment of buses.

The performance of the proposed system is evaluated through a case study conducted in a laboratory environment using CORSIM. Results show that the models developed at the three levels are superior to the signal control implemented in the field, and the alternatives that use the off-line MULTIBAND model for progression for all traffic scenarios. Robust optimization was highly effective in reducing control delays, stop times, queues, and bus delays, and increasing throughput and speeds, when traffic volumes were high. The model that integrated bus priority with robust arterial signal control produced the most reductions in bus delays while not causing significant delays to automobiles.

ROBUST OPTIMIZATION MODEL FOR BUS PRIORITY CONTROL UNDER ARTERIAL PROGRESSION

by

Meenakshy Vasudevan

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2005

Advisory Committee:

Professor Gang-Len Chang, Chair
Professor Bruce L. Golden
Professor Ali Haghani
Dr. Henry C. Lieu
Professor Paul M. Schonfeld

DEDICATION

To my husband, Venkat
and
to the memory of my parents

ACKNOWLEDGEMENTS

I wish to express my sincere appreciation and gratitude to my advisor Professor Gang-Len Chang for his guidance, encouragement and constant support throughout my graduate study. He has been a great source of inspiration to me all through my research years.

I am grateful to Professor Bruce Golden, Professor Ali Haghani, Dr. Henry Lieu, and Professor Paul Schonfeld for serving on my committee and for providing their invaluable comments and suggestions.

I am grateful to Greg Hatcher and Karl Wunderlich for supporting me in this endeavor. I am also deeply grateful to Li Zhang and Juan Morales, without whose support I could not have completed this work.

I am profoundly grateful to my brother, Gopal, and my sister, Lakshmi, who provided me with the love, encouragement and opportunity to pursue my aspirations. I would like to express my deepest gratitude to my husband and friend, Venkat, without whose love, tireless sacrifice and encouragement, I could never have completed this research. Finally, I am extremely grateful to my parents for always having had infinite faith in me. I dedicate this work to their memory.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Major metropolitan areas currently face severe traffic congestion that threatens to be detrimental to the quality of life. Intelligent Transportation Systems (ITS) offer the potential for increasing the operational efficiency and capacity of surface transportation systems. To substantially improve urban traffic conditions, it is essential to implement effective strategies from both demand and supply sides. Encouraging the use of public transit systems is one of the advancing demand-side strategies for relieving urban congestion. Since adaptive signal control is one of the most promising supply side methods for relieving urban traffic congestion, the integration of bus priority with adaptive signal control emerges as a critical task.

Priority treatment of buses is in line with the promotion of public transit systems as it can provide an opportunity to reduce bus travel times and passenger waiting times by minimizing delays at intersections. It will also be beneficial to transit operating agencies since it will enable an extension of a bus route without any increase in travel time, and reduce the fleet size without degrading the service quality. However, recognizing that a preferential treatment of bus users is often at the cost of automobile users, the implementation of such a priority strategy has often faced considerable opposition from traffic engineers, and hence, evolves as a challenging task. A rigorous evaluation of all possible trade-offs and complex interactions between transit and automobile users under different traffic conditions is thus essential for a successful implementation of a bus priority system.

In reviewing the literature it was observed that most existing transit priority systems used strategies such as green extension, red truncation, or special bus phase, without considering the impacts of priority treatment on other traffic users. Since priority was constrained only by minimum and maximum green extension durations, such strategies will inevitably cause undue delay to cross street traffic. Most studies also assumed that there were buses only on the major roads and in the peak direction (Courage *et al*., 1978; Garrow *et al*., 1998; Skabardonis, 2000).

Some bus priority strategies used a strategy known as inhibit, which refrained from giving priority to a bus if it had been granted in the previous phase (Richardson *et al*., 1979; Heydecker, 1983). This was designed to prevent disruption to the cross street traffic. However, it seems unreasonable to ignore the priority demand of a bus based on whether priority was granted in the earlier phase or not. The bus in the earlier phase may have been well ahead of schedule and as such, did not warrant a priority treatment. The bus in the current phase may have a significant schedule delay, but with the inhibit strategy adopted by some priority systems, its delay may increase further.

The review of literature has also shown that bus priority is most effective when it is an integral part of traffic signal operations. Otherwise, optimization of signal timings without taking into account bus demands and progression needs may result in disruption of traffic. Ideally, to contend with all the above mentioned concerns, one may expect to have a system that is capable of reliably predicting traffic patterns in real-time, sufficiently efficient in computing arterial progression bands, and integrating bus priority effectively in the local control operations. Thus, the key issues associated with the successful operations of such a system are: “How to incorporate bus priority in the control process without unduly interrupting signal progression?" and "How to generate sufficiently robust arterial progression bands in real-time that can truthfully reflect variation in traffic volumes and the arriving patterns on each link?" The primary focus of this study is to address these critical issues, and develop a real-time arterial progression system with bus-priority under adaptive control environments.

1.2 RESEARCH OBJECTIVES

The purpose of this study is to develop a real-time arterial control with both signal progression and bus-priority functions. The proposed system will consist of three principal components for providing arterial signal progression in real-time, bus priority operations, and local optimization of signals at each intersection. The proposed control logic will not give unconditional priority to buses and cause excessive delays to cross street traffic. Instead, the signal priority decision will be made based on a performance index, which will analyze the benefits of giving priority treatment. Thus, the study will include the following primary tasks:

1. Review current bus priority strategies, both experimental studies as well as those that have been incorporated in existing systems, and identify their limitations as well as strengths.
2. Investigate key progression control strategies, and adaptive signal control methods for isolated intersections, and identify their strengths and deficiencies to develop an integrated arterial signal control model.
3. Design an integrated model that concurrently provides robust progression along the arterial, and optimizes signals at individual intersections in real-time, by

minimizing delays, stops and queues.

4. Integrate bus priority into the robust signal progression control that provides priority to buses by minimizing not only the commonly used passenger delays, and vehicle delays, but also bus schedule delays while allowing for arterial progression.

1.3 ORGANIZATION OF DISSERTATION

To accomplish the stated research objectives, the dissertation is organized as follows:

Chapter 2 presents a review of literature. The first section in this chapter summarizes some of the bus priority strategies that were either experimental or analytical explorations, or implemented in existing systems. The second section discusses some of the adaptive control systems for isolated intersections, which were helpful in the development of the integrated models. The last section reviews the state-of-the-art research in signal control methods for arterial networks.

Chapter 3 presents the overall design architecture of the proposed bus priority system under robust signal progression. The control architecture is divided into three levels: *network* or *progression control*, *local* or *intersection control*, and *bus priority control*. At the *network control level*, progression is provided along the arterial by maximizing variable bandwidths using real-time traffic data. At the *local control level*, vehicle delays, stop times and queue lengths at each intersection are minimized subject to the bandwidth constraints imposed by the network level. At the *bus priority control level*, whenever a bus is detected and is a candidate for priority it is granted priority based on a performance index that is a function of bus schedule delay, automobile and bus passenger delays, and vehicle delays. The functional requirements of the three levels, and the interrelations between their key component modules are described in this chapter.

Chapter 4 discusses the robust arterial signal control model that serves to provide signal progression and local optimization. At the *progression control level*, the arterial progression model generates optimal cycle lengths to maximize variable bandwidths, using a modified version of the MULTIBAND bandwidth maximization model. Traffic signals at each intersection of the arterial are coordinated by providing progression along the arterial, based on real-time surveillance information and predicted results from an artificial neural network model. Queue clearance and minimum green times are functions of existing queues. Bandwidth maximization is performed once every system cycle length. To eliminate uncertainty in the prediction due to the long time horizon and large system, robust discrete optimization technique is employed at this level. At the *local control level*, the signal operations at each intersection are optimized within the progression bands generated by the *network control level*. This is to ensure that progression is not disrupted along the arterial, while optimizing the phasing plans using a rolling horizon approach. The optimization is done every 2 to 5 seconds over a time horizon equal to the cycle length of the intersection, computed at the *network control level*. The objective of the optimization technique is to minimize a weighted-combination of vehicle queue lengths, delays and stop times, with constraints on the bandwidth and minimum green time. This will ensure that the platoons arriving at an intersection will get the allocated bandwidth generated at the *network control level*, and no movement suffers from excessive delays. The function of each component module and a detailed presentation of the two models are given in this chapter.

Chapter 5 presents the integrated bus priority control model under robust signal progression. The model grants priority based on a performance index, which is a function of the schedule delay of the bus, intersection control delays and delays incurred by automobile and bus passengers, subject to bandwidth and minimum green constraints. Minimum green constraints are imposed for driver safety and to discharge average waiting queues. Bandwidth constraints allow for uninterrupted arterial flow despite a preferential treatment of buses. To function efficiently and effectively in real-time, the proposed model consists of three modules that serve to predict bus-related information, check if the bus can compete for priority treatment, and provide bus priority. The component modules, the optimization problem, and the solution algorithm are described in detail in this chapter.

Chapter 6 focuses on evaluating the performance of the developed system, along with a sensitivity analysis. The incremental benefits of the models developed at each of the three levels are evaluated through a case study conducted using CORSIM in a simulation environment. Six signal control alternatives are compared:

(i) Baseline Model, which uses actuated signal control implemented in the field

(ii) MULTIBAND+, which uses the MULTIBAND method for progression control and a short-term signal optimization model for intersection control

(iii) Model I, which uses the MULTIBAND method for progression control and the proposed signal optimization model for intersection control

(iv) Model II, the proposed arterial signal control model that does not use robust optimization or bus priority functions

(v) Model III, the proposed arterial signal control model which employs robust optimization but does not give bus priority

(vi) Model IV, the proposed arterial signal control model which employs robust optimization and gives bus priority

Finally, Chapter 7 presents the conclusions, and recommendations for future research.

CHAPTER 2: LITERATURE REVIEW

2.1 INTRODUCTION

Over the past several decades, a variety of studies related to bus priority have been conducted through either experimental testing or analytical explorations. Some of the bus priority methods have also been implemented in existing signal systems. Overall, the potential benefits of having properly designed and implemented bus priority strategy have been well justified in these studies. However, it should be recognized that bus priority should not be at the cost of passenger car drivers. When signals along an arterial are optimized to provide progression and reduce delays, the problem becomes complicated when bus priority is also considered, since buses drop out of the progression band when they make passenger stops. Thus, a thorough understanding of all possible interactions between bus and passenger car users under various traffic conditions becomes critical in the development of an effective strategy for real-time, bus priority control under signal progression. Hence, the objective of this chapter is to provide a comprehensive review of existing bus priority strategies, signal optimization methods for isolated intersections and arterial networks, and arterial progression approaches.

This chapter is organized as follows: Section 2.2 presents key literature in bus priority strategies, followed by a discussion of some of the adaptive control systems for isolated intersections in Section 2.3. Section 2.4 reviews the state-of-the-art research in signal control methods for arterial networks. Section 2.5 summarizes the review results and concluding comments.

2.2 BUS PRIORITY STRATEGIES

Considerable research has been performed to evaluate the effectiveness of giving priority to buses at intersections, using computer simulation models and analytical models. The objective of these studies has been to evaluate a priority scheme prior to its implementation. Salter and Shahi (1979) employed computer simulation models to evaluate the effects of bus priority on passenger and vehicle delays, queue lengths and bus travel times, for all types of vehicles. Their purpose was not to propose any priority technique, but to show that giving priority to buses increased the delay of private vehicles, while decreasing bus-delay. However, the authors did not disclose the bus priority strategy adopted in their study.

Vincent, Cooper and Wood (1978) suggested the following four control strategies: priority extension or green extension, priority call or red truncation, compensation green to the cross street if the priority call extended beyond the normal maximum green, and inhibit placed on priority call during cross street green, following a granted priority call. Their simulation studies showed that having only priority extension gave very limited benefits to buses. Combining priority extension and call, and inhibit methods, resulted in larger benefits to buses while producing equally larger losses to non-bus users. Combining all the four control methods produced smaller benefits than with the second strategy, but also lower penalty to other traffic users. It was also superior to the first strategy. The authors concluded that giving an extension usually resulted in a net benefit to all passengers and hence, it was not necessary to have an inhibit function in the control strategy.

Richardson and Ogden (1979) disagreed with the idea proposed by Vincent et al (1978) about the redundancy of having an inhibit function. They developed a methodology for evaluating an active bus priority system. In their priority system, each time a bus was awarded priority, the time lost by the non-bus phase was added to the maximum green time in the next cycle. When the non-bus phases accumulated a large amount of green time that was never repaid, the system refused preemption demand by a bus. The authors concluded that the loss caused to other traffic users, by granting recalls in successive cycles, outweighed the benefit gained by buses.

Courage, Wallace and Wattleworth (1978) investigated three bus priority techniques: reversible exclusive bus lane, bus preemption using traffic signals, and traffic signal progression system. In the first priority technique, the reversible lane was used as an exclusive bus lane during peak periods and as an additional left turn lane during off-peak periods. In the second technique, green was granted to a bus subject to a minimum green constraint and a maximum green extension of 120 seconds. At the end of a bus phase, green was provided to the cross street. In the third priority system, traffic signals were coordinated to provide a progression mode favorable to the peak direction. The paper assumed that this was also the direction of travel of a bus, which is not a reasonable assumption. The three techniques were combined in different ways to generate four stages. It was found that preemption did not decrease the green time of the cross street traffic. When no preemption was given, the green time was about 48.5% for the cross street traffic. For stage 2 (a combination of the first and second techniques), the green time increased up to 53.5%. This was because when a bus cleared the intersection, green was automatically given to the cross street. For stages 3 (a combination of the first and third techniques) and 4 (a combination of all three techniques), the green time for the cross street decreased by about 20%, due to the constraint of coordinated operation for progression.

Seward and Taube (1977) evaluated bus-actuated signal preemption systems based on a revenue/cost ratio. The proposed system consisted of three basic components: bus detection time, expected arrival time of the bus at the intersection, and the satisfaction of minimum and maximum green constraints. If a bus arrived at the end of a green phase, then extension of green was granted provided the maximum green constraint was not violated. If a bus arrived at the start of a red phase, then green was not granted until the minimum green constraint was satisfied. In the revenue/cost ratio, bus-operating cost, bus-passenger cost and automobile passenger cost were considered. The authors concluded for a specific street corridor in Milwaukee that, the revenue/cost ratio was greater than one and hence warranted a bus-preemption strategy. The paper showed how the improved bus travel times due to preemption would enable the reduction in the number of buses without any deterioration in service. In the experiment, only one bus approach was handled. The authors did not discuss how their system would guarantee preemption if buses arrived simultaneously on conflicting approaches.

The findings of Khasnabis, Reddy, and Chaudry (1991) agreed with the experimental studies of Seward and Taube (1977), that preemption results in a reduction of fleet size. Their paper evaluated the feasibility of using signal preemption as a tool for transit demand management using an analytical model. In the proposed procedure, priority was given in the form of green extension, or interruption/ termination of a red phase. They developed a software (PREEMPT) to analyze the travel demand, fare box revenue and operating cost as a result of giving preemption to buses. The authors showed that signal preemption resulted in a reduced fleet size and operating cost. However, as demand increased, the fleet size and consequently, the operating cost, increased. On the other hand, this was offset by an increase in the fare box revenue resulting from higher ridership. Despite the promising results, it should be noted that the authors did not evaluate the effect of preemption on drivers of private vehicles.

Ludwick (1976) compared three bus-preemption algorithms. The first algorithm granted only a green extension when the time saved by passengers in the bus street exceeded the time lost by passengers in the cross street, while the second algorithm granted unconditional preemption by using green extension and red truncation whenever a bus approached an intersection. The third algorithm also gave preemption whenever a bus approached the intersection, but had two priority constraints on: the time limit for the duration of priority, and the minimum time elapsed after a priority ended, before giving the next priority. It was found that the third algorithm was the most promising. The first algorithm did not cause any substantial improvement to travel time. The second algorithm resulted in an improvement of 25% to bus travel time, but caused an increase in cross street travel time by 20%. The third algorithm, with a 10-second limit, produced an improvement of 20% to bus travel time, while causing a 7% increase in cross street travel time. The paper concluded that bus-preemption is most beneficial when frequency of bus service is high and bus stops are located on the far side of the intersection.

Jacobson and Sheffi (1981) developed a stochastic model for total passenger delay. As suggested by Ludwick (1976), their experiments showed that preemption is most beneficial when bus occupancy and flow of buses are high. They also found that the waiting time for cross street traffic decreased due to preemption, because in their model, optimal signal control decision was made to minimize the total person delay.

Along the same lines, Heydecker (1983) stated that the benefits of bus preemption must be weighed with the consequences of a reduction in capacity at the intersection. The author investigated bus priority using the strategies of extension and recall by considering an inhibition rule. An extension was granted in any cycle, provided a minimum green time, for safety reasons, had elapsed. A recall was granted only if none was granted in the previous cycle, as recommended by Richardson and Ogden (1979). This was done to restore the capacity of each traffic stream to the level that existed when there was no priority.

Garrow and Machemehl (1998) examined the effectiveness of transit priority on the Guadalupe - N. Lamar arterial in Austin, Texas, and proposed guidelines for its use. They used TRAF-NETSIM as their evaluation tool. During off-peak periods, they examined the effect of reduced signal cycle lengths and split phasing in conjunction with local transit service, and unconditional priority with express transit service. According to the authors, reducing cycle lengths for transit priority was appealing monetarily since it eliminated the use of sophisticated vehicle detection systems. Their simulation studies showed that reducing cycle lengths was beneficial to buses as it reduced their average travel time. Additionally, short cycle lengths did not penalize cross street vehicles. Split phasing did not greatly reduce the travel time of buses, nor did the cross streets experience any change. Their studies indicated that unconditional transit priority was most effective when the saturation level of cross streets dropped below 0.25. This was because the signal timing would favor the bus approach ensuring that a bus would seldom request unconditional priority. The authors also noted that when unconditional priority technique was used with express bus services, the negative impact on the cross streets was minimal since express bus services use longer headways than local bus services, resulting in fewer priority calls.

During peak periods, the authors examined the effect of transit priority at an isolated intersection and within an arterial network. For isolated intersections, they found that the overall travel time per person was lower for a green extension of 10 seconds than that for 20 seconds. They also observed that near-side bus stops were less beneficial to transit priority than far-side bus stops, since a significant portion of the green extension was utilized for the bus dwell time at near-side bus stops. For an arterial network, they evaluated four scenarios, based on the amount of green extension given along the arterial. Their studies showed that travel time along the arterial for buses was least when priority was given, but this resulted in an increase in the travel time for cross street traffic. They concluded that transit priority was most feasible when transit had a high mode split and during peak periods rather than during off-peak periods.

Skabardonis (2000) proposed passive and active transit priority techniques for arterials with coordinated traffic signals. Passive priority strategies included development of signal-timing plans that favored buses along the arterials. This was achieved by modifying the signal plans of TRANSYT-7F by coding the bus movements as separate links. Weighting factors for delays and stops for the bus links were specified so that the signal optimizer would favor the transit vehicles over the rest of the traffic. Experiments showed that fixed time plans favoring buses were most effective when the bus volumes were high and bus arrival times were highly predictable. However, with increase in the uncertainty of arrival times due to dwell time at bus stops, the benefits of priority treatment decreased substantially. The main criteria for the development of active priority strategies included the availability of spare green time in the cycle length, progression at downstream intersections, and schedule adherence. The TRANSYT-7F model was used to generate signal-timing plans for the baseline traffic scenario. Intersections were selected for priority treatment based on the spare green time and the bus arrival times at the intersection. The splits and offsets were optimized again using TRANSYT-7F for those intersections, by assigning a zero weighting factor for the auto links and the maximum weighting factor of 10000 for the bus links. The signal settings for the rest of the intersections were fixed. Tests showed a reduction in delays and stops for buses and an improvement to their speeds. However, it produced excessive queues on the cross streets, resulting in increased delays. Moreover, although bus progression was one of the criteria in the development of the strategy, it was not observed. Buses that were granted priority at one intersection joined the end of the queue at the downstream intersection. The proposed priority methods did not consider transit vehicles on cross streets.

All the aforementioned studies evaluated various bus priority techniques to demonstrate the benefit of granting preemption. Bus preemption caused a reduction in total delay to passengers and fleet size, which resulted in a reduction in the operating cost. Although the findings were promising, they were only experimental studies. There are a few bus priority strategies that have been introduced within existing signal systems such as UTCS/BPS (MacGowan and Fullerton, 1979), SCRAM (Cornwell, 1986), PRODYN (Henry et al, 1983), UTOPIA (Mauro and Taranto, 1989), SCOOT (Hounsell et al, 1996) and the system proposed by McGinley and Stolz (1985). Bus priority strategies are designed to increase passenger throughput as well as decrease delay, and thereby improve the traffic system's control benefit. However, instead of evaluating the consequences of bus priority on other drivers, some of the bus preemption systems gave absolute priority to buses, by using pre-specified strategies such as phase extension, phase early start, special bus phase, and interruption of a red phase.

Signal Coordination of Regional Areas in Melbourne (SCRAM; Cornwell, 1986) and the tram priority system proposed by McGinley and Stolz (1985) were two major projects to coordinate signals and introduce tram priority at intersections in Melbourne, Australia. SCRAM utilized an extended version of SCATS to facilitate tram priority. In SCRAM, tram priority was provided based on two techniques - passive priority and active priority. Passive priority used historical data on tram behavior. Priority was given using strategies such as, reduced cycle time by considering shorter cycles, a green time biased towards a tram approach, a special phase design biased towards tram movement, and linking for tram progression by setting offsets recognizing the lower travel speed. In active priority approach, priority was given only when a tram was detected. This approach used methods such as phase extension, phase early start, special tram phase, phase suppression, and window stretching. SCRAM could provide priority at any point in the signal cycle and had the facility to transfer time between phases.

In the system proposed by McGinley and Stolz, tram priority was achieved by an active priority technique - flexible window stretching. It was the same technique as adopted in SCRAM, but in SCRAM, the overall efficiency was reduced. In SCRAM, the length of the early start, when a bus was detected during a red phase, was a constant. This resulted in wasted early start priority. Whereas in the priority system proposed by

McGinley and Stolz, the tram priority phase was canceled the moment the tram cleared the intersection. Furthermore, unlike SCRAM, their system was also able to handle situations where more than one tram demanded preemption. However, in both systems it was observed that when a tram was detected signals were adjusted without evaluating the effect on the entire system.

The initial objective of UTCS/BPS (MacGowan and Fullerton, 1979), which is part of the UTCS family, was to give priority to a bus at an intersection, without causing delay to traffic on the cross street. However, their approach was very conservative to result in a substantial reduction of delay to buses. Hence, they adopted a strategy of giving preemption to buses that were detected by bus loop detectors, using strategies such as green time extension, or red phase termination. If a bus approached during a red phase, then it was granted a green provided the minimum green of the cross street had been reached. If a bus arrived during the end of a green phase, then green time was held until the bus cleared the intersection.

In the aforementioned systems, preemption was given with constraints only on minimum and maximum green, but without considering the effect of preemption on other drivers. In contrast to these priority systems, PRODYN, UTOPIA and SCOOT evaluated the impact on other traffic before giving priority to buses. PRODYN (Henry *et al.*, 1983) was a pioneering research in computing in real-time the best acyclic signal plans. PRODYN converted the signal optimization problem for a traffic network into several smaller problems. Each smaller problem was assigned to an intersection, which was solved by dynamic programming approach. The global problem was then solved for the network using an iterative procedure of first optimizing at the lower intersection level, and then transferring the signal decision to the upper level, where the supervisor of the network simulated the network with the initial signal decision and found the best control. The optimization function was the sum of delay over a time horizon of 15 intervals, each of duration 5 seconds, with constraints on maximum and minimum green. The original version of PRODYN gave priority to transit vehicles by treating them as several private vehicles. Recently PRODYN was modified to explicitly model transit operations.

Urban Traffic OPtimization by Integrated Automation, (UTOPIA; Mauro and Taranto, 1989) made possible the absolute priority to public transport vehicles at intersections and the actual improvement of private vehicles' mobility in all traffic conditions. Following the same technique as adopted in PRODYN, the problem was decomposed into smaller subproblems belonging to two different classes - the intersection level and the area level. At the intersection level, signal control decision was made based on predicted arrival times of vehicles at the stop lines, by optimizing a function of weighted sums of time lost by vehicles, vehicle stops, maximum queue length for every link of the intersection, and time lost by the public vehicles. The optimization was done over a time horizon of 120 seconds and was repeated every 6 seconds. However, the decision was carried out for only 6 seconds. The area level updated the weights in the optimization function. The weights were obtained by optimizing over a time horizon of 30 minutes the total travel time spent by private vehicles in crossing the area. The UTOPIA system was implemented in Turin, Italy.

The SCOOT system (Bowen *et al*., 1994; Hounsell *et al*., 1996) granted priority to a bus based on a user specified intersection degree of saturation to avoid excessive delays to other traffic. SCOOT did not predict the dwell time of a bus; instead, a bus competed for priority only after it cleared a bus stop. Priority treatment was considered only for buses that were behind schedule in the form of extensions or recalls. When a bus arrived at the end of a green, the current green was extended beyond its scheduled length to allow a bus to clear the stop line if the degree of saturation at the intersection was below a threshold value. When a bus arrived during a red phase, green was recalled if the degree of saturation was below a specified threshold for the intersection. Field evaluations showed bus delay savings of 5 to 10 seconds with no delay to the rest of the traffic.

Yagar and Han (1993, 1994) proposed a procedure, Signal Priority Procedure for Optimization in Real-Time (SPPORT), for determining real-time signal plans that considered transit preemption based on the delay caused to all passengers. It is an acyclic transit priority model that used a rule-based optimization process to generate candidate signal timing plans for different levels of transit priority. The decision to grant a transit priority was made by computing the immediate benefit and the future benefit for each phase. The immediate benefit was defined as the sum of the priorities of all the active requests that would be served by switching to that phase. The future cost was defined as the sum of all priorities of all the inactive requests to which service would be delayed by switching to that phase due to minimum green and inter-green constraints. The phase sequences were evaluated based on a performance index, which was the total passenger delay for the projection period. Projection period was defined as the time-period for which the future traffic information was available. The signal-timing plan that resulted in the least delay was selected.

Real-time signal optimization software has been developed based on the SPPORT model (Conrad and Yagar, 1998). Experiments proved that SPPORT afforded priority to transit vehicles with minimum disruption to other traffic. However, it has yet to be implemented in the field. The delay equation considered the same waiting time for all approaches, which is not rational. The occupancy of vehicles in queue in all approaches was multiplied by the time interval to give the delay. Additionally, the authors did not discuss how to handle situations when a bus was ahead of schedule. A bus that is ahead of schedule does not warrant a priority over other passenger cars. However, if it is far behind its schedule, its weighting factor should be high. Hence, these factors should also be considered while developing a delay model in order to obtain the best signal-timing plan.

Chang, Vasudevan and Su (1995) explored the advantages of integrating bus priority with adaptive signal for isolated intersections. Similar to SPPORT, their model evaluated the effect of bus priority on other traffic users. The decision to grant priority was made every second based on a performance index, a weighted combination of passenger delay, vehicle delay, and bus schedule delay. The objective function was evaluated for each possible signal state and the most beneficial signal state was selected as the optimal decision. A minimum green constraint was imposed for each phase, for driver safety and for discharging the average queue. It should be noted that unlike other bus priority strategies, this study considered the effect of bus priority on bus schedule delay. However, the signal decision was made based on the system performance for the next second, without considering the future consequences.

Recently, two studies (Duerr, 2000; Balke *et al*., 2000) have been mentioned in the literature that considered bus priority along arterial networks. They addressed the need for minimum disruption to progression along the arterial while granting priority to buses. Balke *et al*. (2000) proposed a method that gave priority to a bus without losing signal coordination along an arterial. The design objectives were to provide priority without disrupting progression and significantly altering the normal sequencing and duration of the non-coordinated phases. Priority was given only to those buses that were in need of priority, based on some user-defined criteria. The system architecture had four modules: (1) Arrival time prediction module, (2) Priority Assessment Module, (3) Strategy Selection Module, and (4) Strategy Implementation Module. When a bus was detected by a GPS automatic tracking/locating system, the arrival time prediction module used the GPS data and historical information to predict the arrival time of a bus at the bus stop, the dwell time of the bus, and the arrival time of the bus at the stop line. In the priority assessment module, an evaluation was performed whether a bus was a candidate for priority treatment or not. In their study, this criterion was schedule adherence. When a bus was late by 5 minutes, it was flagged for priority treatment. Other buses were not granted priority. In the strategy selection module, a bus priority strategy from strategies such as phase extension, early return, and special bus phase insertion was selected. The criteria for the strategy selection was that none of the phases in the background-timing plan would be skipped and each phase, if activated, would be provided at least its minimum green and clearance interval. Once the strategy was selected, it was implemented in the strategy implementation module.

Experiments showed that there was only a minimal increase to the overall system delay as long as the volume to capacity ratio was below 0.95. The cross streets experienced substantial increase in delays. This is to be expected since this method did not evaluate the impact on passenger car users when priority is given to buses. Priority was granted as long as the minimum green and clearance time constraints were satisfied, and none of the phases were skipped.

Duerr (2000) proposed a bus priority control system called DARVIN (Dynamic Allocation of Right-of-Way for Transit Vehicles In Urban Networks) to improve bus progression in mixed traffic while optimizing overall performance of the network. The three main objectives of the system were to reduce travel times for transit vehicles, maximize schedule adherence, and minimize disruption to general traffic. A microscopic traffic simulation tool was first designed since none of the existing simulation tools satisfied DARVIN's requirements for performing simulations sufficiently fast, representing buses individually on a microscopic level, and modeling interactions between cars and buses in mixed traffic. A genetic algorithm approach was used to optimize the signals by minimizing a weighted-combination of delays and stops. Experiments showed that DARVIN produced significant reductions to overall passenger delays.

2.3 SIGNAL CONTROL FOR ISOLATED INTERSECTIONS

In this section, adaptive control systems whose objective is to provide an optimum signal control decision for a specified control time-period for an isolated intersection will be reviewed. These systems do not aim at obtaining an optimal cycle length, but try to determine the optimal signal control. There are two types of approaches for optimization of the signal control: (1) binary approach and (2) rolling horizon approach. In the binary approach, optimal decision is made based on a trade-off analysis that considers the advantages of extending a current green against the disadvantages of terminating it. In the rolling horizon approach, the signal control is optimized within a rolling horizon framework using either a sequential search method or dynamic programming.

2.3.1 Binary Approach

Miller's algorithm (Miller, 1963) is the forerunner for such algorithms where the signal decision is based on the trade-off between extending green and terminating it. It is modeled for modifying signal-timing plans in small time steps of 2 seconds. In each time step, a delay-based optimization function is used to evaluate the net benefit of putting off the termination by 2, 4, 6, 8 and 10 seconds, as opposed to terminating the green immediately. If the net benefit is positive, the current green is extended for another two seconds, and the control function is used again to make a decision for the next two seconds. However, in this algorithm, the length of the next red phase is assumed to be equal to the length of the last red phase. This assumption is not reasonable since it does not take into account the number of vehicles queuing on the red phase.

In Traffic Optimization Logic (TOL) proposed by Bang (1976), an optimal control decision to extend green by h seconds or terminate it immediately, is made by a trade-off in the net benefit calculated from the system objective function. Unlike Miller's algorithm, TOL designates the next red duration as the green duration needed by the current red phase to discharge its queuing vehicles. Moreover, TOL considers only two options of adjusting the current green - terminating the green immediately or extending by *h* sec.

Microprocessor Optimized Vehicle Actuation (MOVA), proposed by Vincent and Peirce (1988) makes a decision as to how long the green signal is to be displayed based on traffic flow and queue information got from detectors. In MOVA, each approach is given enough green to safely discharge queuing vehicles between the stop line and the upstream detector located at 40 meters. After the end of minimum green, green is held until at least one lane of an approach is discharging at below saturation rate. At the end of saturation flow, the merit of extending and terminating green is made based on a performance index which is a combination of vehicle delay and vehicle stops.

The binary approaches consider only a very short projection horizon and hence, do not guarantee global optimal control. Consequently, their performance in light traffic conditions is worse than actuated control systems. Some of the adaptive control systems that make use of the rolling horizon approach will be discussed in Section 2.3.2.

2.3.2 Rolling Horizon Approach

The rolling horizon approach was first proposed by Gartner (1983) to solve the traffic signal optimization problem. Optimization Policies for Adaptive Control (OPAC; Gartner, 1983) is an on-line signal optimization algorithm designed to optimize the performance of individual traffic signals. It uses a sequential search approach to identify the optimal signal-timing plan. In this approach, the optimal signal switching sequence is determined for a future time-period, known as the optimization stage or horizon, which is divided into a number of time intervals. Feasible signal sequences for all time intervals are identified, and the switching sequence that minimizes the performance index is chosen as the optimal plan. The performance index is a weighted-sum of delays and stops. The optimal switching plan is implemented only for the first several time intervals for which traffic flow data is directly available from the detectors. However, this version of OPAC, known as ROPAC, can only handle two phases for each intersection. ROPAC calls for at least one phase change per horizon length. This requirement is flawed since even if there are no calls from the side street, ROPAC will service the side street once per horizon and cause unwarranted delays to the major street.

OPAC was modified to handle dual-ring, eight phase operations (Gartner *et al*., 1991). Both versions of OPAC have been extensively tested. Field tests show that the OPAC system is most effective when the volume is high. It has now been enhanced to optimize traffic signals for a coordinated system of intersections (Pooran *et al*., 1999). This will be discussed in Section 2.4. The OPAC models use an optimal sequential constrained search method, which does not guarantee global optimality.

Controlled Optimization of Phases (COP; Sen *et al*., 1997) optimizes traffic signal control at an intersection using dynamic programming by minimizing a performance index, which is a combination of delays, stops, and queue lengths, with constraints on minimum green and clearance time interval for safety. The model also allows phase sequencing and phase skipping. The study indicates that the COP algorithm is highly responsive to vehicle arrivals. This is to be expected since the three performance indices are functions of vehicle arrivals. Hence, if the vehicle prediction is incorrect, the signal plan will be incorrect. However, the COP algorithm makes use of a simplistic, short-term traffic flow prediction model to predict vehicle arrivals using detector data, signal plan and queue lengths at the upstream intersection. Moreover, the prediction model uses constant free-flow speeds to estimate the arrival time of vehicles at the intersection.

2.4 SIGNAL CONTROL FOR ARTERIAL STREETS

The most widely used signal control methods for arterial networks make use of the following two approaches: (1) minimization of a disutility function, such as delays and stops, and (2) maximization of progression bands. In Section2.4.1, the core concepts of key disutility-based methods will be discussed. Section 2.4.2 will review the most widely used progression strategies.

2.4.1 Disutility-Based Arterial Signal Control Methods

TRANSYT-7F (Wallace et al, 1984) belongs to the first set of signal optimization strategies that minimize a disutility function. It is one of the most widely used signal optimization software for arterial networks. It is a macroscopic tool, for determining phase splits and offsets by minimizing a performance index, a weighted-combination of delays, stops and queue spill back. It uses a gradient search technique as the optimization procedure. The offsets and green times are changed in an iterative manner, and the performance index is computed and compared with the previous value to check for any improvement. Although TRANSYT-7F is able to optimize signal settings for arterial networks and is one of the most widely used models for minimizing delays, it has a few deficiencies. The TRANSYT-7F model uses a gradient search technique to optimize the green time, but this optimization technique does not guarantee a global optimum solution. It requires a signal-timing plan as an initial solution and the quality of the final solution depends on the initial solution due to the nature of gradient search technique. Moreover, it does not optimize the sequence of left-turn and through phases. Additionally, TRANSYT-7F assumes that the average flow demand remains constant.

The SCOOT (Split, Cycle, and Offset Optimization Technique) System was developed by TRRL. It evolved from TRANSYT and is known as the on-line version of TRANSYT (Martin and Hockaday, 1995). SCOOT adjusts green splits, offsets and cycle times, based on traffic demand obtained from detectors, to minimize delays and stops. The on-line traffic model in SCOOT stores information from detectors in the form of cyclic flow profiles and uses cruise speeds and saturation flow rates to predict queues, stops and delays. The underlying philosophy of the system's optimization model is to minimize disruptive alterations to the signal timing plans. Hence, the three optimizers that control green splits, offsets, and cycle times make changes of 4 seconds in multiples of one second. The model computes the degree of saturation for all intersections and determines the critical intersection. The cycle optimizer computes the optimal cycle time, once every five minutes, for the critical intersection so that it operates at or below 90 percent degree of saturation. It assigns the optimum cycle time to all intersections. However, it also has the capability of assigning half the cycle time to an intersection provided progression can be maintained. The split optimizer, before every phase change, minimizes the degree of saturation on all approaches to an intersection by determining whether to increase or decrease the green times by up to 4 seconds or leave them unaltered. The offset optimizer minimizes the performance index, a function of delays and stops, at each intersection once every cycle and adjusts the offset by up to 4 seconds.

Field evaluations in the United Kingdom show that SCOOT performs best when traffic flows are heavy and close to saturation. Moreover, since it does not require long-term traffic data, it is not sensitive to detector failures. However, it does not optimize phase sequences, and its global optimality is not guaranteed.

The Sydney Coordinated Adaptive Traffic System (SCATS) was developed by the Australian Roads and Traffic Authority of New South Wales in 1970's (Lowrie, 1982, Grubba *et al*., 1993). It is a computer based on-line traffic signal control system. The principal objectives of the system include minimizing overall stops, delays and travel times; maximizing throughput; and minimizing the formation of queues, based on information from vehicle detectors located on each lane. SCATS has a two-level coordinated hierarchical architecture for strategic and tactical control. At the strategic control level, optimum cycle times, offsets and splits are determined for the area based on the traffic conditions. The optimum cycle time is determined as the one that maintains the highest degree of saturation and the optimization takes place every cycle. The cycle time is increased or decreased to maintain a degree of saturation of 0.9 on the lane with the greatest degree of saturation. The optimum phase splits are determined as those that maintain equal degrees of saturation on competing approaches, thereby minimizing delays. Free flow speed and degree of saturation are used to determine optimum offsets as those that minimize stops and delays for traffic flows along the arterial.

At the tactical control level, the signal operations at each intersection are modified within the constraints imposed by the strategic control level. Such modifications to signal operations include phase skip, phase termination, or phase extension to its maximum value. However, the main phase cannot be skipped or terminated early. Any time saved during the cycle due to phase skip or phase termination is used by subsequent phases or added to the main phase to maintain the system cycle length.

It should be noted that in SCATS a common cycle length is assigned to all intersections in the arterial even if some of the intersections are not operating at optimality. The SCATS system has been implemented in many cities throughout the world. A study conducted by the Australian Research Board indicated that SCATS showed no significant reduction in travel times as compared to TRANSYT. However, there was a large reduction in the number of stops (Luk *et al*., 1982). SCATS has recently been implemented in Oakland County, Michigan. A study (Taylor, 1997) conducted to determine the change in travel time concluded that the travel time improved for both directions for both the peak and non-peak periods. The savings ranged from 6.6% to 31.8%.

OPAC has now been extended for coordinated signals operation. The coordinated OPAC real-time adaptive control system was first implemented on a network of 16 intersections on the Reston Parkway in Reston, Virginia (Pooran, 1999). This version is known as the Virtual Fixed Cycle OPAC (VFC-OPAC) because from cycle to cycle the yield point or local cycle reference point is allowed to range about the fixed yield points. The VFC-OPAC algorithm determines phase splits by optimizing a performance index, a weighted-combination of delays and stops, with constraints on minimum and maximum green, cycle lengths and offsets. It determines the virtual cycle as the cycle length of the critical intersection. The critical intersection is the one with the highest delay. VFC-OPAC increases or decreases the cycle length only if the previous two decisions regarding the cycle length support the decision. The model allows the virtual cycle to be changed only by 1 second in one cycle. At the end of each cycle, offset optimization is performed based on data from adjacent intersections.

OPAC allows phase skipping in the absence of demand. Under oversaturated conditions, the objective is to maximize the throughput. The model provides maximum green to the congested phases with constraints on the cycle length. Thus, green is given to the phase that allows the largest number of vehicles to leave the link. At the same time, the upstream signal is set so that the smallest number of vehicles enter the link. One advantage of OPAC is that if there is a communications failure between adjacent intersections, the isolated intersection version of OPAC will kick in. However, like the isolated intersection version, VFC-OPAC does not optimize phase sequences. Field evaluations of VFC-OPAC on the Reston-Parkway (Owen et al, 1999) showed that OPAC resulted in an increase of 10% in the average travel time and an increase in the stopped delay. It was concluded that OPAC was unable to determine optimal cycle lengths, splits and progression.

Disutility-based approaches consider the actual traffic demand and try to minimize delays and stops for the entire arterial system. Partial progression may be achieved by these methods. However, progression bands produced by these models may be neither continuous nor wide. Hence, in the next section state-of-the-art studies on arterial progression methods will be discussed.

2.4.2 Arterial Progression Methods

The most widely used bandwidth-based methods are PASSER II and MAXBAND. Most bandwidth maximization approaches try to set the signal timings so that the number of vehicles that can traverse the arterial in both directions without stopping is maximized.

PASSER II (Progression Analysis and Signal System Evaluation Routine) is a macroscopic deterministic optimization model first developed by Messer *et al*. (1973). It has since undergone a series of enhancements and is now known as PASSER II-90 (1991). It uses an iterative gradient search method to determine the best combination of phase sequences and cycle lengths for maximizing progression along the arterial in both directions. The PASSER II model combines Brook's Interference Algorithm with Little's Optimized Unequal Bandwidth Equation. The model first determines the optimal demand-to-capacity ratios and uses them to determine the splits. Cycle lengths, phase patterns, and offsets are varied to determine the optimal set of signal control settings, which minimize the total interference to the progression bands. It can also select multiple phase sequences.

MAXBAND, an arterial progression model that maximizes bandwidths was developed by Little, Kelson, and Gartner (1981). MAXBAND incorporates variable speeds and cycle lengths to accommodate the fact that a small change in the speed or cycle length can result in different signal timings and bandwidths. It is able to calculate the optimal cycle time, offsets, speeds, and phase sequences to maximize the weighted-combination of bandwidths. It computes the green splits from traffic volumes and capacities, using Webster's equation. MAXBAND employs the queue clearance time to permit secondary flows accumulated during red to discharge before platoon arrivals. The optimization problem is formulated as a mixed-integer linear programming problem and is coded with FORTRAN IV.

It should be noted that both MAXBAND and PASSER II make the assumption that there is no or very few traffic entering the arterial from the side street or making a left turn from the arterial. The two arterial progression strategies do not consider the actual traffic volumes, and hence are insensitive to variations in traffic conditions. Both models consider uniform bandwidths, and a constant directional ratio between the inbound and outbound bandwidths. As is observed from field evaluations that a single parameter cannot represent the directional volume ratio for the entire arterial, both progression models cannot guarantee an optimal bandwidth progression for varying traffic flow patterns.

MULTIBAND, a variable-bandwidth arterial progression scheme, was developed by Gartner et al (1990) as an extension of the MAXBAND model. Unlike MAXBAND and PASSER II, MULTIBAND assumes variable bandwidths. This allows for variation in traffic volumes along the arterial due to traffic turning into or out of side streets. It incorporates link-specific bandwidths, inbound to outbound bandwidth ratios, volumes, and speeds in its maximization problem. The objective function addresses the problem of varying traffic flow patterns. Moreover, each bandwidth is given a unique weight corresponding to its contribution to the optimization function. The weights are defined to be a function of the link-specific volume and capacity. Mixed-integer linear programming technique is used to solve the problem.

The aforementioned three bandwidth-maximization programs allow the user to specify a queue clearance time, mainly used to handle the secondary flows before the through platoons arrive at the intersection. The queue clearance time is pre-specified by the user. This value cannot be a constant, but varies with the queue length at each intersection. Moreover, all of the above mentioned progression strategies are off-line methods and use historical data.

Unlike, the previously mentioned off-line progression strategies, REALBAND (Dell 'Olmo and Mirchandani, 1995) provides signal coordination along arterial networks in real-time. It forms variable progression bands based on actual traffic data, and the widths as well as the speed of a progression band are determined by optimizing a network-wide performance index. The REALBAND model first identifies platoons by filtering detector data and predicts their movement in the network. Each platoon is assumed to have a different speed. It then sets the traffic signals to give suitable green time to the platoons to optimize a performance index such as stops and delays. When there is a conflicting demand, one platoon is given priority over the other, or one of the platoons is split to maximize the performance index. The optimization is done over a time horizon and a decision tree of various candidate decisions is formed. The branch that results in the least cost is chosen as the optimal signal plan. However, REALBAND requires an initial solution of phase timings, which is used as an upper bound for the performance. Furthermore, it was found to be effective for light to moderate traffic conditions but not for over-saturated conditions.

As reviewed in the previous two sections, optimization of signal timings may result in disruption of traffic, and arterial progression does not reduce delays on side streets nor does it provide a system-wide reduction in delays. When arterial progression is considered along with signal optimization, there is a smoother flow of traffic along with the advantages of optimization, such as reduction in delays and stops. Hence, in the next section some studies that have been conducted to examine the benefits of combining the best features of both approaches will be examined.

2.4.3 Integration of Disutility and Progression Methods for Arterial Control

A number of studies have been conducted to combine the disutility-based approaches and the bandwidth maximization approaches. In order to overcome some of the deficiencies in the TRANSYT model, Cohen (1983) combined the progression opportunities offered by the MAXBAND model and the delay minimization of the TRANSYT-7F model. In this paper, the author uses the MAXBAND model to generate a bandwidth solution, which is used as the starting input for TRANSYT-7F. The results of Cohen's study indicate that while this approach improves the signal-timing plans, progression is not guaranteed.

Cohen and Liu (1986) developed a modified version of TRANSYT-7F called the TRANSYT-7F(C). In the constrained model, the system delay is minimized while maintaining a two-way progression. Analogous to TRANSYT-7F, this model changes the offsets and green times by small amounts, and compares the performance index with the previous value. However, after each shift a check is performed to prevent any red time, including the dual left-turn green time, from encroaching upon the through band generated by the MAXBAND model. Thus, the signal timings generated by TRANSYT are constrained by the progression bands of the MAXBAND model. Their study shows that the unlike the bandwidth starting approach, the constrained model is able to provide progression. This approach does not discriminate against side streets, because the slack green time can be made available to the side street if the performance index can be improved by offering the additional green to side streets.

Wallace and Courage (1982) proposed a new arterial progression approach that maximizes progression opportunities (PROS) along the arterial signal system. Progression opportunity is defined as “the opportunity, presented at a given traffic signal and at a given point in time, to travel through a downstream signal without stopping.” Thus, it is the number of successive green signals that a driver faces when travelling without stopping, at the desired speed. This approach performs only as well as any other maximal bandwidth approach since actual traffic demand is not considered. Hence, the authors further expanded the PROS method to include simultaneous minimization of delays and stops. The new objective function is defined as PROS/PI, where PI is the performance index of the TRANSYT model. Experiments show that the PROS/PI method results in significant reductions in main street delays and total delays while providing progression. However, by including the PROS in TRANSYT’s objective function, the PROS/PI approach discriminates against the side streets (Cohen and Liu, 1986).

Chang *et al*. (1985) modified PASSER II to minimize arterial delay while preserving the maximum bandwidth solution. In their approach, the offsets at an intersection are adjusted within the slack time to minimize delays, based on the traffic flows from the upstream intersection. Thus, the model maintains the two-way progression band while reducing delays. Unlike the PROS/PI method, this approach does not discriminate against the side streets since additional green from the side streets is not taken to reduce delays. Instead, only the offsets are adjusted within the slack time, when available. However, it should be noted that the green times are fixed and cannot be adjusted to reduce delays.

The real-time, hierarchical, distributed traffic signal control (RHODES; Head *et al*., 1992, 1998) uses the same methodology of hierarchical distribution of control as used by PRODYN and UTOPIA. It decomposes the signal control problem into three hierarchical levels: (1) Intersection Control, (2) Network Control and (3) Network Loading. At the highest level, the network loading level predicts the travel demand over long periods of time, which is used to determine the sizes of the future platoons. At the middle level, the network control level determines the coordination constraints for each intersection, by predicting the platoon flows. The Network Control Logic is composed of two components, the Platoon Flow Prediction Logic and the Network Optimization Logic. APRES-NET (Dell 'Olmo and Mirchandani, 1994) is the macroscopic simulator used to propagate platoons through the network. It uses detector actuations to estimate the arrival time of a vehicle at the intersection, signal timing plans to determine when the vehicle will leave the intersection and the turn probability to assign a direction of travel. The Network Optimization logic is based on the REALBAND method described in Section 2.4.2. At the intersection control level, the optimal signal timings and phase sequences are computed with the COP method described in Section 2.3.2. The local optimization performed by COP is constrained by the coordination bands of the Network Control Level. However, the Intersection Control Level can adjust the start and end time of a phase based on the platoon flow on that link.

Bus priority has been implemented at isolated intersections, but due to the adverse impacts on private vehicles in a coordinated system, has faced considerable resistance from traffic engineers. The literature has shown that there are very few implementations of bus priority systems on arterial networks. Thus, it becomes essential to develop a bus priority system under arterial signal progression that improves the performance of bus operations, while not being detrimental to other drivers.

2.5 SUMMARY

In review of the existing literature, it is clear that considerable progress has been made in the past. Early experimental studies used absolute preemption techniques with constraints only on minimum green and maximum green extension times. In order to mitigate the negative impact to other drivers, recent bus priority approaches have tried to integrate bus priority within signal control operations. Some of the studies also indicated that a bus should not be given priority unless it was behind schedule. However, it can be observed that none of the bus priority strategies discussed how to handle situations when buses arrived simultaneously on conflicting approaches. It was assumed in most priority techniques that there were buses only on major roads and in the peak direction. More importantly, none of the existing bus priority systems and signal optimization/progression systems dealt with uncertainty in traffic flow data. Despite advances in surveillance technology, there is considerable amount of uncertainty in traffic behavior. It is likely that input errors, from any source, to the optimal control process may significantly mislead the search for an optimal signal control solution, thereby yielding undesired suboptimal solutions.

A rigorous evaluation of all possible trade-off and complex interactions between transit and automobile users under various traffic conditions is necessary for any strategy development and potential applications. Thus in order to effectively execute bus priority along an arterial, the control algorithm should address the following major issues:

- Provide progression control in real-time: Signal control plans are effective only if they are responsive to real-time fluctuations in traffic.
- Project evolution of traffic flow along the arterial: To ensure signal progression along the arterial, it is necessary to reliably project traffic conditions in advance along the arterial, rather than use historical data or simple mathematical models.
- Provide progression bands that are fully utilized: Since arterial progression control generally consists of multiple intersections and is over a long period of time, it is difficult to pre-specify a queue clearance time that fully reflects the arriving traffic dynamics.
- Generate progression bands of variable bandwidths: This is designed to accommodate variations in traffic volumes along the arterial due to turning flows at intersections.
- Provide progression along the arterial without being disadvantageous to cross-street traffic: To prevent excessive delays to side streets, minimum green constraints should be imposed to guarantee the side-street traffic a green time, if there is a demand.
- Provide signal optimization with traffic signal coordination: This ensures a smoother flow of traffic while reducing system-wide delays.

- Incorporate bus priority as an integral part of the traffic signal operations to reduce bus delays: Rather than use pre-specified strategies, such as phase extension, phase early start, and/or a special bus phase, priority decision should be made based on a performance index, which is a function of vehicles delays, bus schedule delays, and delays caused to users of buses as well as automobiles.
- Provide bus priority without causing excessive delays to other traffic users: To prevent excessive delays, minimum green constraints should be imposed based on both the existing traffic conditions and driver safety.

To address the above concerns, in this research a real-time arterial traffic control system that provides bus priority while simultaneously maximizing progression bandwidths and optimizing the signal timing plans at each intersection along the arterial is developed. The next chapter presents the architecture of the arterial traffic control system.

CHAPTER 3: ARCHITECTURE OF THE ROBUST OPTIMIZATION MODEL FOR BUS PRIORITY UNDER SIGNAL PROGRESSION

3.1 INTRODUCTION

As is well recognized, a preferential treatment of bus users is often at the cost of passenger car drivers. Hence, the implementation of bus priority control has become a very sensitive issue. The review of literature has shown that bus priority is most effective when it is an integral part of traffic signal operations. However, to ensure an effective bus priority operation at the arterial level, one needs to contend with the following two critical issues: (1) How to maximize the signal progression band and ensure that it will not be interrupted unduly for bus priority? (2) How to take into account inevitable uncertainties associated with input data for maximizing progression and determining bus priority eligibility? The latter issue is especially complex for an arterial of multiple intersections as the spatial evolution of traffic platoons, and the variation in volumes and turning flows in each link are extremely difficult to predict accurately and reliably. Depending on the level of deviation between the actual and predicted information, a system incapable of tackling such an issue may yield non-optimal control strategies that offer neither signal progression nor justifiable bus priority for field implementation.

In view of the aforementioned concerns, this study intends to present a multi-level design for a real-time robust arterial traffic control system that provides bus priority while simultaneously maximizing progression bandwidths and optimizing signal timing plans at each intersection along the arterial. A detailed description of the system architecture, along with the functional requirements of each principal component, and their interrelations will be described in Section 3.2, followed by the summary in Section 3.3.

3.2 SYSTEM ARCHITECTURE

The proposed system architecture aims to provide efficient bus priority control and minimize overall system delay under a real-time signal progression environment. The control architecture is divided into three levels: *network* or *progression control*, *local* or *intersection control*, and *bus priority control*. At the *network control level*, progression is provided along the arterial by maximizing variable bandwidths using real-time traffic data. At the *local control level*, vehicle delays, stop times and queue lengths at each intersection are minimized subject to the bandwidth constraints imposed by the network level. At the *bus priority control level*, whenever a bus is detected and is a candidate for priority it is granted priority based on a performance index that is a function of bus schedule delay, automobile and bus passenger delays, and vehicle delays. The interrelations between these three levels of operations and their principal components are illustrated in Figure 3.1, and presented in detail below.

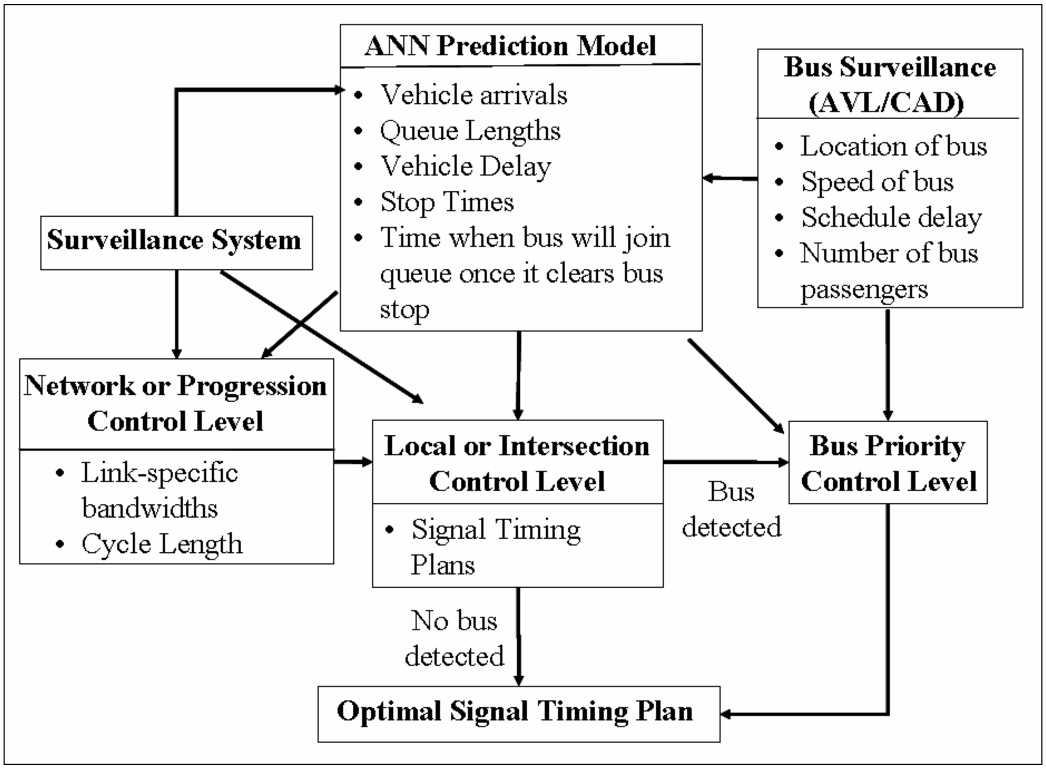

Figure 3.1 Design Architecture of the Bus Priority Control System under Signal Progression

3.2.1 Network or Progression Control Level

At the *progression* or *network control level*, traffic signals at each intersection of the arterial will be coordinated by providing progression along the arterial, based on real-time surveillance information and predicted results from an artificial neural network (ANN) model. The arterial progression optimization model will generate optimal cycle lengths, platoon speeds, and offsets to maximize variable bandwidths in both inbound and outbound directions along the arterial.

To assume the primary role in the entire control process, the proposed progression control model will have the following functions:

- **Capable of providing progression control in real-time**

To implement bus priority in real-time, the employed signal progression model should also function in real-time. Unfortunately, most existing arterial progression strategies such as MAXBAND, PASSER II, and MULTIBAND are off-line in nature, using mainly historical average volumes for computing coordinated signal offsets. For real-time applications, signal control plans are effective only if they are responsive to real-time fluctuations in traffic behavior. A through band often cannot be fully utilized due to inherent discrepancies between the historical traffic data and actual platoon patterns. Hence, the optimization model for signal progression should be capable of using on-line surveillance information and generate real-time control plans.

- **Capable of projecting evolution of traffic flow along the arterial**

To ensure signal progression along the arterial, it is necessary to be able to reliably project the traffic conditions in advance along the arterial. Most existing signal progression methods have relied only on historical data (e.g., MAXBAND, MULTIBAND, etc.) or simple mathematical models (REALBAND, RHODES) to approximate the variability in traffic demand and driver behavior over time. However, the discrepancies between actual and approximated traffic patterns often exceed the acceptable level, and may significantly mislead the search for optimal progression signal settings. Thus, despite the complexity in projecting the time-varying traffic patterns, it is essential for a real-time progression model to have a function that can capture the dynamic evolution and interactions among traffic demand, signal operations and driver behavior. This research proposes the use of artificial neural networks, as it is ideally suited for recognizing patterns or forecasting data where conventional approaches, such as regression-based methods, may fail. Note that with regression-based methods one has to make assumptions about the functional form of the traffic measurements. Moreover, regression models cannot suitably capture the dynamics of traffic conditions (Chien and Liu, 2002). The advantages of using ANN include its fault tolerance, its ability to learn and adapt, and its massively parallel-distributed structure. An ANN model can overlook small distortions to the input data. Hence in this study, artificial neural networks will be used to predict the queue lengths at each intersection along the arterial for the entire time horizon. Real-time data from the surveillance system, as shown in Figure 3.1, will be used as input for training the ANN model for arterial flow prediction. The ANN model will also be capable of predicting vehicle arrivals at each intersection.

- **Capable of providing a progression band that is fully utilized**

A progression band can be fully realized only if the platoon flow is not hampered by any queue at the intersection along the arterial. Existing progression strategies usually handle this by using a queue clearance time to clear the queue in advance of the arriving platoons. However, models such as MAXBAND and MULTIBAND can only use a queue clearance time, pre-specified by the user. Since arterial progression control generally consists of multiple intersections and is over a long period of time (typically equal to the system cycle length), it is not possible to pre-specify a queue clearance time that fully reflects the arriving traffic dynamics. Hence, the proposed model will use a queue clearance time computed with time-varying queue that is to be provided by the embedded ANN arterial flow prediction module.

- **Capable of generating progression bands of variable bandwidths**

Since it is not realistic to assume a uniform platoon or constant volume moving along the entire arterial, the proposed progression control model will be capable of generating variable bandwidths for each link and in each direction. This is designed to accommodate variations in traffic volumes along the arterial due to turning flows at intersections. To further capture the impact of turning flows on the progression band, each road section (link) in the proposed model will be assigned a corresponding directional ratio since it is unrealistic to assume a constant directional volume ratio throughout the arterial. A link-specific directional ratio can reflect the diversity in traffic. Moreover, the optimization function will be to maximize variable bandwidths. Each bandwidth will be given a unique weight corresponding to its contribution to the optimization function. The weights will be a function of the link-specific volume.

- **Capable of providing progression along the arterial without being disadvantageous to cross-street traffic**

Most progression models provide a good progression along the arterial but at the expense of side-street traffic. To prevent excessive delays to side streets and to allow sufficient time for pedestrians to safely cross the arterial streets, minimum green constraints should be imposed. This will guarantee the side-street traffic a green time equal to at least the minimum green time, if there is a demand. Such an essential feature will be included in the proposed model.

- **Capable of accommodating the uncertainty in traffic flow data through robust optimization**

As is well recognized, progression models try to platoon traffic along the arterial by providing a through band. Most existing models use either historical data or simple mathematical computations to estimate traffic volume. However, neither historical data nor data computed using deterministic or stochastic optimization models can fully capture the uncertainties in traffic flow data. As signal progression is typically provided over a long period and for multiple intersections, the discrepancy between the projected and actual arriving flows could be so significant as to mislead the optimal search process and yield ineffective progression offsets. This is especially critical when historical or partial real-time data are employed for signal optimization. Hence, the proposed model will mitigate the uncertainty by using robust discrete optimization, which is a comprehensive mathematical programming framework for handling uncertainties in decision making. It will be discussed in detail in the next chapter.

With the aforementioned functional requirements in mind, this research intends to take full advantage of available work, and take it as the basis for necessary enhancements. In the review of literature, it has been found that the MULTIBAND model, unlike the other bandwidth maximization strategies, incorporates variable bandwidths. It allows for variation in traffic volumes along the arterial due to traffic turning into or out of side streets. The model assigns a directional ratio to each road section (link) of the arterial, instead of assuming a constant volume ratio. Moreover, each bandwidth is given a unique weight corresponding to its contribution to the optimization function. The weights are defined to be a function of the link-specific volume. Thus, MULTIBAND emerges as one of the most superior bandwidth maximization strategies.

However, it should be noted that MULTIBAND is still an off-line strategy. It does not consider traffic flows in real-time and therefore, is not applicable in real-time traffic control problems. It allows the user to pre-specify a queue clearance time, which is used to handle the secondary flows before the through platoon arrives at the intersection. This value in a real-world system should neither be a constant nor pre-specified, but vary with the existing queue length, since with such reliable time-varying information the optimization control model can clear the actual queues ahead of the platoon traffic and provide a meaningful progression. Hence, a modified version of the MULTIBAND model, which satisfies all the functional requirements, will be used as the arterial progression optimization control.

To provide all essential functions, the proposed *progression control level* will consist of the following modules: (1) the arterial flow prediction module, (2) the signal progression module, and (3) the robust discrete optimization module. As shown in Figure 3.1, the arterial flow prediction module functions to predict the queue lengths, vehicle arrivals and link volumes for the entire cycle length using a neural network model.

The signal progression module serves to maximize a set of variable bandwidths. In progression control, uniform or variable bandwidths are maximized by defining bandwidth ratio constraints, loop integer constraints, speed and cycle length constraints, and directional interference constraints. In this study, the optimization function will maximize weighted variable bandwidths subject to not only the aforementioned constraints, but also additional real-time operational constraints for queue clearance and minimum green time. A complete set of constraints for progression optimization are defined as follows:

1. **Directional Interference Constraints**: Directional interference constraints are applied to prevent progression bands from traveling during the red time.
2. **Loop Integer Constraints**: Loop integer constraints result from the fact that the intersections of the arterial are synchronized, i.e., they are operating on a common cycle length.
3. **Bandwidth Ratio Constraints**: These constraints avoid restricting the larger band to a fixed ratio once the smaller band has reached its maximum. Bandwidth ratios are taken to be the ratios between the inbound volumes and the outbound volumes along the arterial.
4. **Cycle Length Constraints**: An optimal common cycle time is obtained by constraining it within pre-defined upper and lower limits. All intersections operate at the common cycle length.
5. **Speed Constraints**: These constraints are applied to prevent wide fluctuations in speed from one link to the next.
6. **Queue Clearance Constraints**: These constraints are applied to define queue clearance time to be the time to clear the queue at the beginning of the green time. The queue clearance time is the time taken to clear the queue ahead of the platoon to prevent the through bands from not being fully utilized. Some progression methods include a queue clearance time in their optimization, but it is pre-specified and does not represent the actual queue.
7. **Minimum Green Constraints**: Minimum green constraints are applied to

prevent side streets from suffering when arterial volumes are high and for pedestrian safety. The minimum green time will be a function of the queue at the intersection.

The third module uses the robust discrete optimization technique to contend with input data uncertainties, and to generate optimal link-specific bandwidths, cycle lengths, and offsets, over all realizable traffic scenarios.

Note that signal progression does not necessarily lower delays. In the proposed system, progression control is integrated with optimization of signal timings by using bandwidths generated from the progression strategy as constraints to the signal optimization problem, which is solved at the *intersection control level*. The *network control level* will pass the following information to the *intersection control level*:

- progression bands for each link and in each direction
- system cycle length

3.2.2 Local or Intersection Control Level

As noted previously, the *progression control level* is responsible for generating a robust solution to the variable bandwidth problem over the system cycle length. However, as progression control does not guarantee reduction in delays, it is essential for the system to provide local or intersection control for minimizing local delays and accommodate traffic variations due to modeling of multiple signals. At the *intersection control level*, optimal signal timing plans for each intersection will be generated under the bandwidth constraints of the *network or progression control level*. In summary, the proposed *intersection control level* shall have the following functions:

- **Capable of providing real-time adaptive signal control with the ability to skip phases**

For real-time applications, the intersection optimization should adapt to traffic variations in real-time. The optimization at each intersection will be performed using adaptive signal control since neither pre-timed nor actuated signal control are suitable for simultaneous demand-responsiveness and coordination functions. The objective function will be to minimize a weighted combination of vehicle queues, delays and stop times over a time horizon equal to the cycle length. The optimization technique will have the ability to skip phases in the event that there is no vehicle demand and will be solved using dynamic programming approach.

- **Capable of performing optimization to ensure both traffic progression and bus priority**

The objective of the proposed system is to provide bus priority under signal progression, without causing delays to automobiles on the arterial and on side streets. Signal optimization models reduce system delays, but may cause interrupted traffic flow. In contrast, progression methods provide a smooth flow of traffic along the arterial, but do not reduce delays or queues, especially for side streets. Hence, it is proposed that the bus priority control system should take advantage of the strengths of both these methods. The signal timing plans will be optimized for each intersection on the arterial, within the bandwidth constraints determined by the *progression control level*. The signal phasing plans will be generated so that the total green time allocated for the through movement along the arterial in the outbound (inbound) direction is greater than or equal to the bandwidth for the outbound (inbound) direction.

- **Capable of predicting real-time lane specific queue lengths for each control time step over the entire cycle length**

Some existing models, such as RHODES, tend to capture complex traffic patterns with a simple static queue estimation model which is a function of uniform vehicle speeds, vehicle actuations at the upstream detector and uniform queue dispersion rate. The level of accuracy achieved by such a model is not sufficient for a large system-wide optimal progression control. Hence, artificial neural networks will be used to predict queue lengths and vehicle arrivals that are lane and link specific, as well-trained ANN models can easily adapt to changes in real-time.

To perform the above functions, the *intersection control level* will have the following three modules: (1) the traffic state prediction module, (2) the signal state estimation module, and (3) the signal optimization module. The traffic state prediction module serves to generate vehicle queue lengths, control delays and stop times for each control period using an artificial neural network prediction model. The signal state estimation module monitors the signal state, the elapsed green time for each phase and the elapsed arterial through green, and estimates the minimum green duration. The signal optimization module handles the optimization of the signal by minimizing vehicle queue lengths, delays and stop times subject to bandwidth and minimum green constraints. Constraints associated with this level are defined as follows:

1. **Bandwidth Constraints**: These constraints ensure that progression is maintained along the arterial under optimized local signal control plans. Thus, the platoons traveling along the arterial will receive the bandwidth made available from the *progression control level*. Thus reduction in delays will be achieved while

guaranteeing signal progression.

2. **Minimum Green Constraints**: The minimum green constraints are applied for driver and pedestrian safety and to discharge the average waiting queue during each phase. The minimum green is a function of the real-time queue predicted by the ANN model.

The intersection optimization function will be solved every 2 to 5 seconds. Once the cycle length is nearly completed, the *progression control level* will solve the bandwidth optimization problem again. Note that the bandwidth and signal optimization problems are solved simultaneously, thereby reducing the computation time, and ensuring signal control in real-time. When a bus is detected by the bus surveillance system the system passes control to the next level, the *bus priority control level*.

3.2.3 Bus Priority Control Level

The *progression control* and the *intersection control levels* serve mainly to provide arterial signal operations. No priority treatment for buses is considered at these two levels. When the surveillance system identifies a bus, the *intersection level* will pass the control to the *bus priority level*, wherein bus priority operation will be activated. To effectively interact with the two higher-level operations, the proposed *bus priority level* shall be capable of providing the following functions:

- **Capable of providing bus priority while operating within a signal system**

As reflected in the literature, to reduce bus delays while not at the expense of other traffic users, bus priority needs to be embedded in the adaptive signal control functions. Rather than using pre-specified strategies, such as phase extension, phase early start, and/or a special bus phase, as in most conventional transit priority systems, a minimum green constraint will be imposed whether there is a bus demand or not. Such a constraint will be based on both the existing traffic conditions and driver and pedestrian safety. Moreover, a bus will enter the demand only when detected by a bus-detector or surveillance system. Until then, it will be treated as an automobile. A bus that is detected will not automatically warrant priority. Priority treatment will be based on a set of rules. The system evaluation will be based on a performance index, which is a weighted combination of automobile and bus passenger delays, intersection control delays, and bus schedule delays.

- **Capable of providing bus priority without discounting signal progression**

In a coordinated signal system, bus priority treatment can cause disruptive service, thereby causing concern among the signal operations community whose main objective is to provide a smooth flow of traffic. Unless bus priority treatment is considered without disrupting arterial progression, promoting use of public transit systems by affording priority will cease to be an attractive way of dealing with congestion. Hence in this model, when a bus is detected and is a candidate for priority, the priority decision will be evaluated based on a performance index, which is a weighted combination of automobile and bus passenger delays, vehicle delays, and bus schedule delays, subject to bandwidth constraints. By including bandwidth constraints, the system will be able to afford bus priority by considering its effect on arterial progression.

- **Capable of considering schedule adherence in the priority decision**

Schedule adherence is important for efficient bus operations, improving service quality, and increasing transit users' satisfaction. This is one of the objectives of the

Advanced Public Transportation Systems Program. Hence, the model will include a bus schedule delay in the optimization function.

To offer the above functions, the *bus priority control level* has the following three principal components: (1) the bus prediction module, (2) the priority candidacy module, and (3) bus priority optimization module. The bus prediction module obtains real-time bus data from the surveillance system, and employs artificial neural networks to predict the following information:

- Bus arrivals
- Time when a bus will join the queue once it clears the bus stop

The priority candidacy module checks if the bus is a candidate for priority based on a set of rules. If it determines that the bus can compete for priority, the control is passed to the bus priority optimization module, where the signals are optimized by minimizing a performance index, which is a weighted combination of automobile and bus passenger delays, vehicle delays, and bus schedule delays, subject to bandwidth and minimum green constraints, which are defined as follows:

1. **Bandwidth Constraints**: These constraints ensure that progression is maintained along the arterial even when affording priority to buses. Thus, the platoons traveling along the arterial will receive the bandwidth made available from the *progression control level*.
2. **Minimum Green Constraints**: Minimum green constraints are applied to ensure that a preferential treatment of buses is not at the cost of automobiles, and for driver and pedestrian safety. Minimum green is a function of the real-time queue predicted by the ANN model.

3.3 SUMMARY

This chapter presents the architecture of the bus priority control system under real-time signal progression. To contend with all the technical and control issues, the system is divided into three levels, the *network* or the *progression control level*, the *local* or the *intersection control level*, and the *bus priority control level*. The *network level* provides arterial progression using robust optimization technique; the *intersection control level* is responsible for optimizing signals within the progression constraints; and the *bus priority level* functions to process priority treatment for buses when detected. The functional requirements for each of these three levels were presented. Chapter 4 presents the arterial signal control problem, which is jointly solved at the *progression control level* and the *intersection control level*.

CHAPTER 4: A REAL-TIME ROBUST OPTIMIZATION MODEL FOR SIGNAL PROGRESSION

4.1 INTRODUCTION

This chapter presents the arterial signal control model that is decomposed into two levels: (1) *network* or *progression control level* and (2) *local* or *intersection control level*. At the *network control level*, signal progression is provided and optimal bandwidths and cycle lengths are obtained using a modified version of the MULTIBAND model. At the *local control level*, signal timing plans constrained within the progression bands, are generated by minimizing a weighted-combination of vehicle queue lengths, vehicle delays and stop times.

The MULTIBAND method will first be evaluated in Section 4.2. The proposed arterial progression method will be discussed in Section 4.3, followed by a description of the proposed intersection optimization model in Section 4.4. The summary will be presented in Section 4.5.

4.2 MULTIBAND PROGRESSION CONTROL METHOD

The objective of signal progression models is to maintain a continuous flow of traffic along the arterial, by coordinating the signal timings of adjacent intersections. To ensure a green band for platoons traveling along the arterial, the MULTIBAND model is subjected to directional interference constraints, bandwidth ratio constraints, cycle length and speed constraints, and loop integer constraints.

Unlike other bandwidth maximization models, the MULTIBAND model has many unique features, such as variable bandwidths, disaggregate bandwidth weighting, and advance queue clearance time. However, it has a few limitations that make it less appealing for real-time applications. The following are some of its main deficiencies:

- **The progression band is not fully utilized.**

In MULTIBAND, to fully utilize the maximum bandwidth, the through band is advanced by an amount equal to the queue clearance to clear existing queues due to turning traffic, ahead of the arriving platoons. However, if the queue clearance time is not a function of the actual queues at the intersections and is assigned arbitrarily by users, the effectiveness of including it to maintain a smooth flow of arterial traffic will be lost due to insufficient or excessive queue clearance time.

If the assigned value is lower than what is actually required, queues will not be cleared before the arrival of platoons, causing queuing of arterial traffic and reduction of progression. If the assigned value is higher than what is required, arterial traffic on the upstream intersections and side street traffic will experience unnecessary delays due to wasted green time. MULTIBAND provides progression control by platooning traffic. However, if queued vehicles impede the lead vehicle in the platoon, the advantages of platooning will be lost since the following vehicles in the platoon are also impeded.

- **Progression control is at the expense of side-street traffic.**

Progression control methods provide a good arterial solution but at the expense of side-street traffic. The objective of the MULTIBAND model is to maximize the bandwidths based on platoon or through flows and saturation flows on the arterial streets. Side-street volumes are included in the formulation by assigning variable bandwidths for each link to account for traffic turning into and out of side streets. However, the variable bandwidth is not a function of the actual turning volume. Hence, when arterial volumes are high, the model will try to maximize the through bands by disregarding the side-street traffic, thereby causing excessive delays to side streets. This limitation cannot be overlooked especially in the context of bus priority, since it is unrealistic to assume that a bus will travel only along the arterial and not on the side streets. Including a minimum green for side-street traffic can eliminate this drawback, since a minimum green constraint will ensure that the side-street traffic is granted a green time that is at least equal to the minimum green, if there is a demand. To provide a progression control that is not detrimental to side-street traffic, it is necessary to impose a minimum green constraint, but the MULTIBAND model does not have such a constraint.

- **The progression model is deficient in real-time applications.**

MULTIBAND performs off-line optimization of variable bandwidths using historical data. Signal progression or optimization models that are off-line and use historical data are deficient in real-time applications. For real-time applications, signal control strategies should be able to adapt to variations in traffic patterns and provide dynamic or real-time timing plans. If a through band is generated using historical data, it will not be fully utilized and signal coordination will be meaningless, due to inaccurate assumptions of traffic demand and platoon arrivals. Fluctuations in traffic demand with time should be considered for providing progression control in real-time. This can only be achieved by using real-time surveillance data, rather than historical data.

- **The progression model does not handle any uncertainty in the data.**

MULTIBAND optimizes variable bandwidths using historical data. Hence, if the outcome is different from the assumed traffic scenario, the progression band that is generated will not be fully realized. In the signal progression problem, as optimization is done only once during the control period, which is typically equal to the system cycle length, and since multiple intersections are modeled, the assumption of having perfectly reliable input data can result in suboptimal or ineffective solutions. Even if the progression method was modified to consider real-time surveillance data, uncertainties of traffic evolution patterns cannot be accounted for over a long control period, as deterministic and stochastic optimization models fail to optimize over all plausible scenarios. It should be noted that none of the existing progression based methods and signal optimization models deal with this problem.

4.3 NETWORK OR PROGRESSION CONTROL LEVEL

This section presents the proposed arterial progression control strategy. To overcome the deficiencies in the MULTIBAND model noted in the previous section, the arterial progression control model is designed to have the following features:

- **The queue clearance time is a function of the actual queue length.**

To prevent the interruption of a smooth arterial flow and for full utilization of the bandwidth, the queue clearance time is a function of the actual time-varying queue at each intersection along the arterial, in the outbound and inbound directions. Thus, the queue clearance time is defined to include: (1) the start-up lost time, (2) the time to clear vehicles that were not able to clear during the previous green, (3) the time to clear vehicles that arrived at the intersection during red time, and (4) the time to clear vehicles that arrive ahead of the platoon, during green.

To determine the queue clearance time, the development of a model for predicting vehicle arrivals and queue lengths at each intersection along the arterial is essential. The model should be capable of taking into account the evolution of traffic flows along the arterial, fluctuations in traffic demand, driver responses to each signal state, and the signal timing plans at each intersection, for each control time interval over the entire time horizon. Existing traffic flow models are not able to capture the complexity of the problem. It has been shown that for intersection queue prediction, neural network prediction models are able to provide an acceptable prediction accuracy of more than 90% (Chang and Su, 1995). Hence in this study, queue lengths at each control time interval for the entire time horizon are obtained using a neural network prediction model, discussed in detail in Section 4.3.2.

- **A minimum green constraint is included.**

A minimum green criterion is imposed on the green time for each phase. This ensures that side-streets are not penalized when the arterial street volumes are high, and pedestrians have sufficient time to cross the arterial street. Thus, progression control is provided along the arterial without causing significant delays to side-street traffic. The minimum green time is a function of predicted queue lengths.

- **The progression control model is solved for the required performance in real-time applications.**

For real-time application, the bandwidth maximization problem is solved once during every system cycle length, which ranges from 60 to 200 seconds, using real-time surveillance information. The challenging task of an accurate prediction of time-dependent arterial flow patterns is accomplished using a neural network prediction model. Data from the surveillance systems are used as inputs for training the ANN model.

- **The progression control model is capable of generating robust optimal results in response to data uncertainties or variations.**

The objective of progression based strategies is to platoon traffic along the arterial by providing a through band. For real-time applications, any uncertainty in the traffic flow data will inevitably result in the through band not being fully realized, due to discrepancies between actual and predicted traffic patterns, including, arterial and side-street traffic demand, platoon arrivals, and queue clearance time. Despite using real-time surveillance information in the proposed model, the decision may not be optimal since optimization is being performed only once during the cycle length. To minimize the prediction errors with respect to traffic demand and traffic flow operations, the study employs the method of robust discrete optimization to obtain an optimal solution. The robust optimization technique will be discussed in Section 4.3.3.

At the *progression* or *network control level*, the system will provide coordination of traffic signals along the arterial by optimizing variable bandwidths. Optimization is done over a time horizon equal to the system cycle length. The proposed system at the *network control level* will have three modules - (1) the signal progression module, (2) the arterial flow prediction module, and (3) the robust discrete optimization module. These three modules are discussed below in detail.

Consider an arterial of n signalized intersections. Figure 4.1 displays the time-space diagram showing the progression of green bands in the inbound and outbound directions. Table 4.1 gives the notations used in the formulation of the problem. In the proposed model, all intersections are operating on a common cycle length, C, and the signal timing variables are expressed in units of the common cycle length. Each link (section of road between two intersections) has a unique band of width b_k assigned to it.

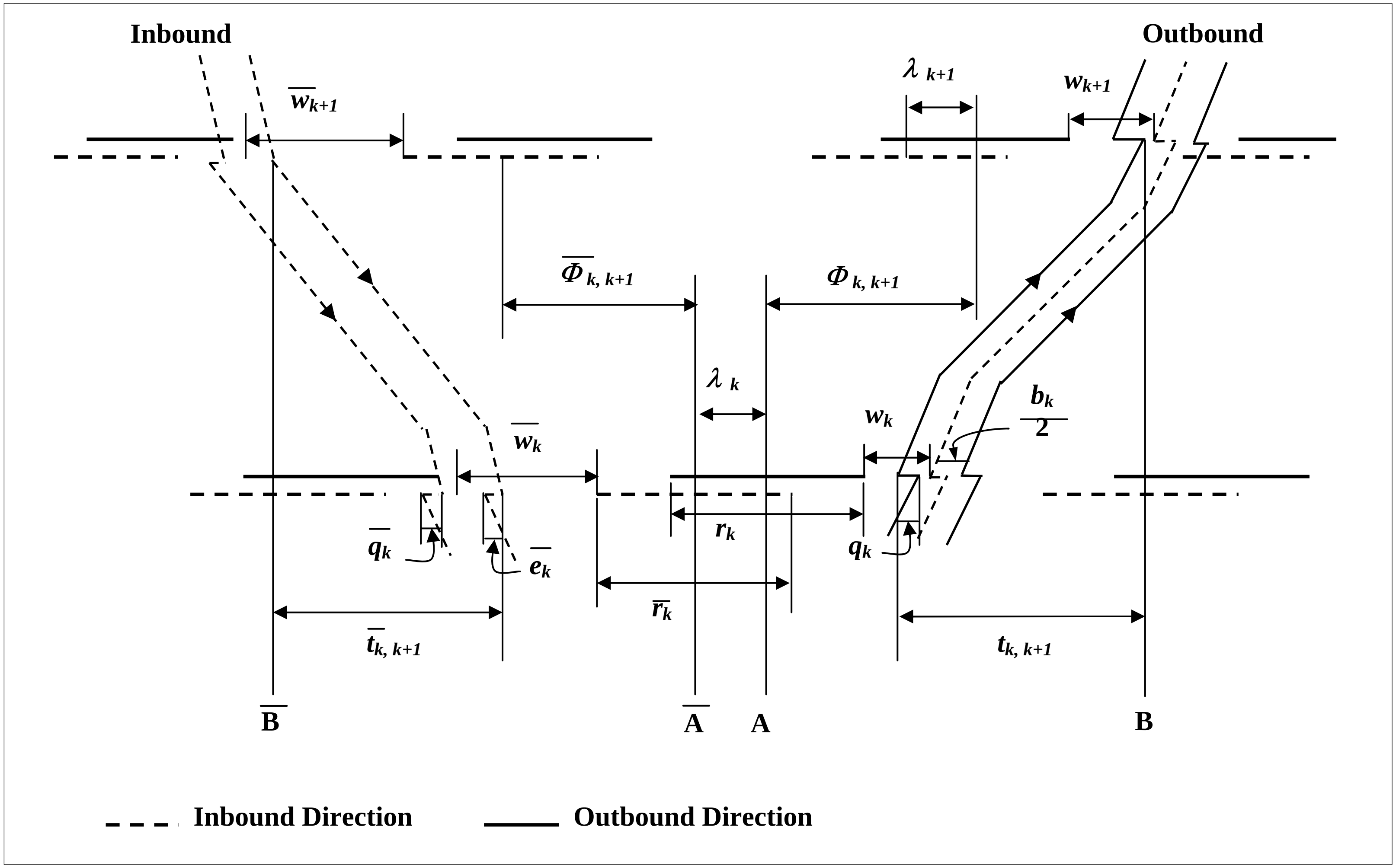

Figure 4.1 Time-Space Diagram for the Inbound and Outbound Green Bands

Table 4.1 Notations Used in the Network Control Level

C	:	System cycle length (seconds)
$\Phi_{k,k+1}$ ($\overline{\Phi}_{k,k+1}$)	:	Internode offsets measured in units of the cycle length, C, is the time from center of an outbound (inbound) red at intersection k to the center of a particular outbound (inbound) red at intersection $k+1$
r_k ($\overline{r}_k$)	:	Outbound (inbound) red time at intersection k, measured in units of cycle length, C
λ_k	:	Intranode offset at intersection k, measured in units of cycle length, C, is the time from center of $\overline{r}_k$ to nearest center of r_k (positive if center of r_k is to the right of center of $\overline{r}_k$)
$m(k, k+1)$	:	Loop integer variable, it is an integral number of the system cycle length, C, for intersections, k and $k+1$
b_k ($\overline{b}_k$)	:	Outbound (inbound) bandwidth at intersection k, in units of cycle length, C
w_k ($\overline{w}_k$)	:	Time from end of red at intersection k to centerline of outbound green band, in units of the cycle length, C (Time from centerline of inbound green band to beginning of red at intersection k, in units of the cycle length, C)
$\overline{e}_k$	:	Shift in the outside edge of the inbound platoon at intersection k (Note that that $\overline{e}_k \geq 0$)
g_k ($\overline{g}_k$)	:	Outbound (inbound) green time for through traffic at intersection k, in units of cycle length, C
l_k ($\overline{l}_k$)	:	Outbound (inbound) green time for left-turn traffic at intersection k, in units of cycle length, C
R_k	:	Common red time in both directions to provide for cross-street movements at intersection k, in units of cycle length, C
q_k ($\overline{q}_k$)	:	Queue clearance time, an advance of the outbound (inbound) bandwidth at intersection k, to clear up queued traffic before arrival of main-street platoon
$t_{k,k+1}$ ($\overline{t}_{k,k+1}$)	:	Outbound travel time at intersection k, i.e., travel time from intersection k to $k+1$, in units of cycle length, C (Inbound travel time at intersection k, i.e., travel time from intersection $k+1$ to k, in units of cycle length, C)
d_k	:	Target ratio of inbound to outbound bandwidths at intersection k
v_k ($\overline{v}_k$)	:	Outbound (inbound) speed at intersection, k (ft/sec)
(v_k^l, v_k^u)	:	Lower and upper limits on the outbound speed at intersection, k (ft/sec)
($\overline{v}_k^l, \overline{v}_k^u$)	:	Lower and upper limits on the inbound speed at intersection, k (ft/sec)
($\delta v_k^l, \delta v_k^u$)	:	Lower and upper limits on the change in the outbound speed at intersection, k (ft/sec)

$(\delta \overline{v}_k^l, \delta \overline{v}_k^u)$	:	Lower and upper limits on the change in the inbound speed at intersection, *k* (ft/sec)
δ_k	:	$\begin{cases} 0 & \text{if inbound left lags at intersection } k \\ 1 & \text{if inbound left leads at intersection } k \end{cases}$
$\overline{\delta}_k$	:	$\begin{cases} 0 & \text{if outbound left lags at intersection } k \\ 1 & \text{if outbound left leads at intersection } k \end{cases}$
$a_k\ (\overline{a}_k)$	:	Outbound (inbound) link-specific weight at intersection, *k*
$V_k\ (\overline{V}_k)$	:	Outbound (inbound) directional volume from (to) intersection *k*; either total volume or through (platoon) volume is used (vph)
$S_k\ (\overline{S}_k)$	:	Outbound (inbound) saturation flow at intersection, *k* (vphpg)
p	:	Exponential power (p = 0, 1, 2, 4)
$S_l^k\ (\overline{S}_l^k)$	:	Outbound (inbound) saturation flow rate for lane *l* at intersection *k* (vphpgpl)
$AL_{k-1,k}$	:	Set of all lanes on the arterial link *(k-1, k)*
SL_k	:	Set of all lanes on the side street links approaching intersection *k*
$Q_l^k(t)(\overline{Q}_l^k(t))$	:	Outbound (inbound) queue length at intersection *k*, at control time interval *t* in lane *l* where $l \in AL_{k-1,k}$ ($l \in AL_{k+1,k}$)
$TQ_l^k(t)(\overline{TQ}_l^k(t))$	:	Outbound (inbound) through queue length at intersection *k* and at control time interval *t* in lane *l* where $l \in AL_{k-1,k}$ ($l \in AL_{k+1,k}$)
$LQ_l^k(t)(\overline{LQ}_l^k(t))$	:	Outbound (inbound) left-turn queue length at intersection *k* and at control time interval *t* in lane *l* where $l \in AL_{k-1,k}$ ($l \in AL_{k+1,k}$)
$SQ_l^k(t)$	:	Side street queue length at intersection *k* and at control time interval *t* in lane *l* where $l \in SL_k$
G_{min}^T	:	Pre-specified minimum green time for through traffic along the arterial (seconds)
G_{min}^L	:	Pre-specified minimum green time for left-turning traffic along the arterial (seconds)
G_{min}^S	:	Pre-specified minimum green time for side street traffic (seconds)
$G_{\min}^{k,T}\ (\overline{G}_{\min}^{k,T})$	:	Minimum green for outbound (inbound) through traffic along the arterial at intersection *k* (seconds)
$G_{\min}^{k,L}\ (\overline{G}_{\min}^{k,L})$	:	Minimum green for outbound (inbound) left-turning traffic along the arterial at intersection *k* (seconds)
$G_{\min}^{k,S}$	:	Minimum green for side street traffic at intersection *k* (seconds)
t_s^p	:	Start-up lost time (seconds)

4.3.1 Signal Progression Module

This module is used to determine variable progression bands. A description of the constraints used in the formulation of the problem is presented first, followed by an in-depth discussion of its modeling logic.

Queue Clearance Time Constraint

Queue clearance time is defined to be the time to clear the queue at the beginning of green time. The queue is a function of: (1) vehicles that were not able to clear during the previous green, (2) vehicles that arrived at the intersection during the red time interval, and (3) vehicles that are not part of the platoon, but arrive ahead of the platoon during the current green. Vehicles that arrive during red time are turning vehicles from the side streets at the upstream intersection. It is not possible to pre-define a queue clearance time, as is done in the MULTIBAND model. In the proposed model, the queue clearance time will also include the start-up lost time, t_s^p.

$$q_k = \frac{t_s^p + \underset{t}{Max}\left\{\underset{l}{Max}\left\{\left(Q_l^k(t)\right)\frac{3600}{S_l^k}\right\}\right\}}{C}, \quad \forall l \in AL_{k-1,k}, t \in [1,T], \forall k \qquad 4.1$$

The queue clearance time is the sum of the following:

- start up lost time due to switching of signals, and
- maximum of the time taken to discharge the queued vehicles over all lanes on each arterial link in the outbound direction and for all control time intervals in the time horizon

A similar equation is also defined for the inbound direction.

$$\bar{q}_k = \frac{t_s^p + \underset{t}{Max}\left\{ \underset{l}{Max}\left\{ \left(\bar{Q}_l^k(t) \right) \frac{3600}{\bar{S}_l^k} \right\} \right\}}{C}, \ \forall l \in AL_{k+1,k}, \ t \in [1,T], \ \forall k \qquad 4.2$$

$Q_l^k(t)$ and $\overline{Q}_l^k(t)$ will be obtained from the ANN prediction model for each control time interval in the time horizon. Note that the queue clearance times in the outbound and inbound directions are defined as units of the system cycle length, *C*.

Minimum Green Constraint

The minimum green constraint is applied to prevent inordinate delays to side streets when arterial volumes are high. This is done by specifying minimum green for arterial links and side street links. It is taken to be the shortest green time during which drivers can react safely to signal changes, while also being sufficiently long for discharging the average queue. It is defined for through and left-turn movements along the arterial, and for side streets as follows:

$$G_{min}^{k,T} = Max\left\{ \left(t_s^p + \frac{3600}{S_l^k} \underset{t}{Max}\left(\underset{l}{Avg}\ TQ_l^k(t) \right) \right), G_{min}^T \right\} \qquad 4.3$$
$$\forall l \in AL_{k-1,k}, \ \forall t \in [1,T], \forall k$$

$$G_{min}^{k,L} = Max\left\{ \left(t_s^p + \frac{3600}{S_l^k} \underset{t}{Max}\left(\underset{l}{Avg}\ LQ_l^k(t) \right) \right), G_{min}^L \right\} \qquad 4.4$$
$$\forall l \in AL_{k-1,k}, \ \forall t \in [1,T], \forall k$$

$$G_{min}^{k,S} = Max\left\{ \left(t_s^p + \frac{3600}{S_l^k} \underset{t}{Max}\left(\underset{l}{Avg}\ SQ_l^k(t) \right) \right), G_{min}^S \right\} \qquad 4.5$$
$$\forall l \in SL_k, \ \forall t \in [1,T], \forall k$$

Equation 4.3 corresponds to the minimum green for through movement along the arterial in the outbound direction. Equation 4.4 is the minimum green for left-turning traffic along the arterial in the outbound direction and equation 4.5 corresponds to the minimum green for side street traffic. Minimum green is the maximum of the following two terms:

- start up lost time due to switching of signals plus the time taken to discharge the maximum of the average queues for all control time intervals in the time horizon
- pre-specified minimum green time to allow drivers to react safely to signal changes and for pedestrians to safely cross the street

The minimum green constraints are defined to ensure that the green time for traffic on arterial and side-street links are longer than the minimum green for the corresponding movements. This will ensure that side-streets do not get unduly short green time. Thus, the minimum green constraints are expressed as follows:

$$g_k \geq \frac{G_{min}^{k,T}}{C} \qquad \forall k \qquad 4.6$$

$$l_k \geq \frac{G_{min}^{k,L}}{C} \qquad \forall k \qquad 4.7$$

$$R_k \geq \frac{G_{min}^{k,S}}{C} \qquad \forall k \qquad 4.8$$

Equation 4.6 implies that the outbound green should be greater than the minimum green for the outbound through traffic along the arterial. Equation 4.7 ensures that the green for left-turn movement along the arterial is greater than the corresponding minimum green. Equation 4.8 implies that the common red in both directions along the arterial is greater than the minimum green for the side streets. The common red is the total green allocated to all side street links approaching intersection k. Similar equations are also defined for the inbound direction.

$$\overline{G}_{min}^{k,T} = Max\left\{ \left(t_s^p + \frac{3600}{\overline{S}_l^k} \underset{t}{Max} \left(\underset{l}{Avg}\ \overline{TQ}_l^k (t) \right) \right), G_{min}^T \right\} \qquad 4.9$$
$$\forall l \in AL_{k+1,k},\ \forall t \in [1,T],\ \forall k$$

$$\overline{G}_{min}^{k,L} = Max\left\{ \left(t_s^p + \frac{3600}{\overline{S}_l^k} \underset{t}{Max} \left(\underset{l}{Avg}\ \overline{LQ}_l^k (t) \right) \right), G_{min}^L \right\} \qquad 4.10$$
$$\forall l \in AL_{k+1,k},\ \forall t \in [1,T],\ \forall k$$

$$\overline{g}_k \geq \frac{\overline{G}_{min}^{k,T}}{C} \qquad \forall k \qquad 4.11$$

$$\overline{l}_k \geq \frac{\overline{G}_{min}^{k,L}}{C} \qquad \forall k \qquad 4.12$$

Bandwidth Maximization Problem

The bandwidth maximization problem is solved subject to the queue clearance time constraint, the minimum green constraint, and all constraints in the MULTIBAND optimization problem. The problem is formulated as follows:

$$Maximize\ B = \frac{1}{(n-1)} \sum_{k=1}^{n-1} (a_k b_k + \overline{a}_k \overline{b}_k) \qquad 4.13$$

subject to:

$$(1-d_k)\overline{b}_k \geq (1-d_k) d_k b_k \qquad \forall k \qquad 4.14$$

$$\frac{b_k}{2} \leq w_k \leq (1-r_k) - \frac{b_k}{2} \qquad \forall k \qquad 4.15$$

$$\frac{\bar{b}_k}{2} \le \bar{w}_k \le (1-\bar{r}_k) - \frac{\bar{b}_k}{2} \qquad \forall k \qquad 4.16$$

$$\frac{b_k}{2} \le w_{k+1} \le (1-r_{k+1}) - \frac{b_k}{2} \qquad \forall k \qquad 4.17$$

$$\frac{\bar{b}_k}{2} \le \bar{w}_{k+1} \le (1-\bar{r}_{k+1}) - \frac{\bar{b}_k}{2} \qquad \forall k \qquad 4.18$$

$$C_1 \le C \le C_2 \qquad \forall k \qquad 4.19$$

$$v_k^l \le v_k \le v_k^u \qquad \forall k \qquad 4.20$$

$$\bar{v}_k^l \le \bar{v}_k \le \bar{v}_k^u \qquad \forall k \qquad 4.21$$

$$\delta\, v_k^l \le v_{k+1} - v_k \le \delta\, v_k^u \qquad \forall k \qquad 4.22$$

$$\delta\, \bar{v}_k^l \le \bar{v}_{k+1} - \bar{v}_k \le \delta\, \bar{v}_k^u \qquad \forall k \qquad 4.23$$

$$\begin{aligned}
&\left(w_k + \bar{w}_k\right) - \left(w_{k+1} + \bar{w}_{k+1}\right) - \frac{1}{2}\left(b_k + \bar{b}_k\right) + \frac{1}{2}\left(b_{k+1} + \bar{b}_{k+1}\right) \\
&+ \left(t_{k,k+1} + \bar{t}_{k,k+1}\right) + \left(\delta_k l_k - \bar{\delta}_k \overline{l_k}\right) - \left(\delta_{k+1} l_{k+1} - \bar{\delta}_{k+1} \overline{l_{k+1}}\right) \\
&+ \frac{1}{2}\left(r_k + \bar{r}_k\right) - \frac{1}{2}\left(r_{k+1} + \bar{r}_{k+1}\right) - \left(q_{k+1} + \bar{e}_k\right) = m(k,k+1)
\end{aligned} \qquad 4.24$$

$$q_k = \frac{t_s^p + \underset{t}{Max}\left\{\underset{l}{Max}\left\{\left(Q_l^k\,(t)\right)\frac{3600}{S_l^k}\right\}\right\}}{C}, \ \forall l \in AL_{k-1,k},\ t \in [\,1,T\,], \forall k \qquad 4.25$$

$$\bar{q}_k = \frac{t_s^p + \underset{t}{Max}\left\{\underset{l}{Max}\left\{\left(\bar{Q}_l^k\,(t)\right)\frac{3600}{\bar{S}_l^k}\right\}\right\}}{C}, \ \forall l \in AL_{k+1,k},\ t \in [1,T],\ \forall k \qquad 4.26$$

$$g_k \ge \frac{G_{min}^{k,T}}{C} \qquad \forall k \qquad 4.27$$

$$\bar{g}_k \ge \frac{\bar{G}_{min}^{k,T}}{C} \qquad \forall k \qquad 4.28$$

$$l_k \geq \frac{G_{min}^{k,L}}{C} \qquad \forall k \qquad 4.29$$

$$\bar{l}_k \geq \frac{\overline{G}_{min}^{k,L}}{C} \qquad \forall k \qquad 4.30$$

$$R_k \geq \frac{G_{min}^{k,S}}{C} \qquad \forall k \qquad 4.31$$

m(k, k+1) is an integer;

$C, b_k, \bar{b}_k, r_k, \bar{r}_k, g_k, \bar{g}_k, l_k, \bar{l}_k, R_k, w_k, \bar{w}_k, t_{k,k+1}, \bar{t}_{k,k+1}, q_k, \bar{q}_k, \bar{e}_k,$
$v_k^l, \bar{v}_k^l, v_k^u, \bar{v}_k^u, G_{min}^{k,T}, \overline{G}_{min}^{k,T}, G_{min}^{k,L}, \overline{G}_{min}^{k,L}, G_{min}^{k,S} \geq 0$

4.3.2 Arterial Flow Prediction Module

In this module, real-time data supplied by the detectors are used to develop a neural network model for prediction of vehicle arrivals and queues at each intersection along the arterial and on the side streets for each control time interval during the entire time horizon.

Artificial neural networks are multi-processing computing systems loosely modeled on the human brain, and consist of a number of highly interconnected processing units, called neurons, which can capture patterns and trends in data, that are too complex to be noticed by humans or which cannot be formulated as an algorithmic solution. Each neuron is connected to other neurons by means of communication links and weights are assigned for each link. The weights represent the strength of a transmitted signal in a network. The neuron is the computational unit of the neural network. Each neuron receives input signal from other neurons, aggregates the inputs, transforms it using a mathematical function known as the activation function, and provides a single output which is sent to other neurons, or used as the output from the net. The weights can either be fixed or modified during the training process using a learning algorithm.

McCulloch and Pitts (1943) developed the first neural network model based on their understanding of neurology. Although there was a decline of interest in neural networks in the 1970s, it has since reemerged as a viable tool for solving a wide variety of prediction and pattern recognition problems. There have been numerous research conducted on the use of neural networks in ITS, including automated incident detection and prediction (Cheu *et al.*, 1996, 2004; Jin *et al.*, 2002), predicting arrival time of buses at stops (Jeong *et al.*, 2004), predicting arrival time of trains for preemption strategies at highway-railroad grade crossings (Cho *et al.*, 2003), etc.

A neural network is characterized by the type of connection between its neurons, its learning or training algorithm, and its activation function. There are a variety of neural architectures, but the most commonly used is the multi-layer feedforward network trained by backpropagation. As is inferred from the name, it consists of multiple layers comprising of the input layer, the output layer and the hidden layer. The flow of information in this type of architecture is forward. Thus, training this type of neural network involves the forward flow of the input training pattern, computation using the activation function, backpropagation of the error, and adjustment of the connection weights.

There are several studies cited in the literature that have analyzed the capabilities of multi-layer perceptrons (MLP) in predicting traffic flows. Mark *et al*. (2004) used

MLP with backpropagation of error in predicting travel times. Dougherty and Cobbett (1997) used MLP successfully to predict traffic volumes, speeds and occupancy. Huang *et al*. (2003) found feedforward neural network with backpropagation of error to be an effective tool in predicting traffic speed under adverse weather conditions. Chang and Su (1995) used a multilayer, feedforward neural network model to predict vehicle queues with an accuracy of more than 90%. Hence, in this study the multilayer, feedforward neural network trained by the backpropagation algorithm is used to predict the following traffic variables needed for solving the progression control problem:

- Lane-specific vehicle queues on the arterial main street for use in queue clearance constraints and minimum green constraints
- Lane-specific vehicle queues on side streets for use in minimum green constraint for side street
- Arterial volumes

Figure 4.2 shows the structure of the ANN model. Note that this model also provides the traffic variables necessary for solving the intersection control problem, described in Section 4.4. As this is an exploratory research, the microscopic traffic simulation tool, CORSIM (1999) is used to generate data for training the ANN model. CORSIM can model traffic flow on an integrated system of freeways and arterial networks, and generate a variety of measures of effectiveness. Chapter 6 discusses in detail the development of a neural network prediction model for a case study.

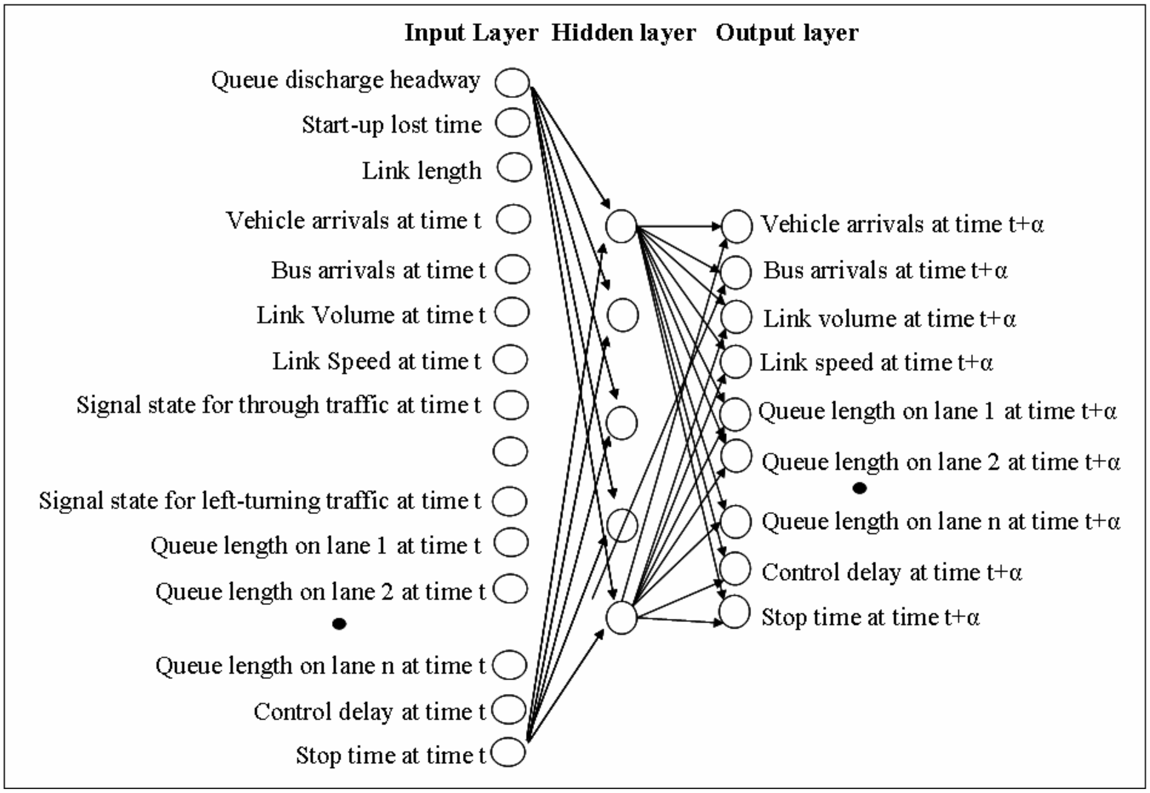

Figure 4.2 Structure of the Arterial Flow ANN Prediction Model

4.3.3 Robust Discrete Optimization Module

In optimizing a proposed system, one may employ either a deterministic or a stochastic method, depending on the nature of the problem and the available information. A deterministic optimization approach typically uses historical data to forecast the future and ignores uncertainty. It uses only the *most likely* data instance to obtain the optimal decision and fails to recognize the possibility of other data instances. In the stochastic optimization approach, the presence of multiple data instances is recognized. However, it assigns a probability to each instance and optimizes the expected performance function over the assumed probability distribution to generate a stochastically optimal decision.

In the signal progression problem, the optimal decision is dependent on the traffic demand, traffic patterns, and driver behavior. Assigning a probability to multiple scenarios of the aforementioned variables is complex and far from trivial, since it is not possible to precisely forecast the occurrence of these time-varying variables over a long period of time and for a large system, particularly due to variations in driver behavior. Moreover, it is also difficult to associate a probability distribution with these variables, due to the presence of uncertain elements such as driver behavior. Additionally, if progression bands are generated using either of the two approaches the system may perform poorly due to the occurrence of a scenario that was not the *most likely* one, or due to an incorrect probability assigned to a data scenario. The *most likely* and the *expected* scenarios are only subsets of the potentially realizable scenarios. Hence, generating progression bands using only one data scenario, either the *most likely*, as is done in the deterministic approach, or the *expected value*, as is done in the stochastic approach, will result in a suboptimal decision if the outcome is substantially different from the assumption. To circumvent the problem and ensure a meaningful progression control that can be fully utilized this study makes use of robust optimization techniques to tackle any uncertainty due to the long-time-horizon prediction in a dynamic system.

Robust optimization generates a decision that best optimizes the performance function under any given data scenario over a pre-specified time horizon. A robust decision is not necessarily the optimal decision, but given any outcome, its performance will be the best among all feasible decisions across all input scenarios. Kouvelis and Yu (1997) defined three robustness criteria in selecting a robust decision: (1) *Absolute Robustness*, (2) *Robust Deviation*, and (3) *Relative Robustness*.

Absolute robust decision best optimizes the objective function among all feasible decisions over all realizable input data scenarios. *Absolute robustness* criterion does not require knowledge of the optimal decision in each scenario. It results in decisions that are conservative in nature since it is based on the assumption that the worst case scenario may occur. However, it is the easiest to compute since it does not require the knowledge of the optimal decision in each scenario (Kouvelis and Yu, 1997).

Robust deviation decision results in the least deviation from optimality, among all feasible decisions over all realizable input data scenarios. *Relative robust* decision is the decision that results in the least percentage deviation from optimality, among all feasible decisions over all realizable input data scenarios. According to Kouvelis and Yu, *robust deviation* and *relative robustness* criteria are less conservative in their decision selection since they determine the decision that keeps its performance as close to the optimal solution for each scenario. Both *robust deviation* and *relative robustness* tend to favor similar decisions. In his study, *robust deviation* will be used as the robustness criterion as it is less conservative than *absolute robustness*.

As robust deviation decision is defined as the one that exhibits the best worst case deviation from optimality, in order to determine the robust decision, X_D, one needs to solve the following problem:

$$z_D = \max_{s \in S}\left(f(X_D, D^s) - f(X_s^*, D^s)\right) = \min_{X \in \cap_{s \in S} F_s}\left(\max_{s \in S}\left(f(X, D^s) - f(X_s^*, D^s)\right)\right) \quad 4.32$$

where:

z_D : Robustness indicator

S : Set of all realizable input data scenarios $(s \in S)$

X_D : Robust deviation decision

D^s : Instance of input data that corresponds to scenario s

X^*_s : Optimal decision for scenario s

X : Set of all decision variables

F_s : Set of all feasible decisions when s is realized

D : Set of input data

Equation 4.32 can be re-written as the following mathematical programming problem:

$$z_D = min \left\{ y \mid f(X, D^s) - z^s \leq y;\ s \in S, X \in \cap_{s \in S} F_s \right\} \quad 4.33$$

where:

y : Worst deviation from optimality, for a given feasible decision, over all realizable input data scenarios

z^s : Value of the objective function at optimality; $f(X^*_s, D^s)$

Since the purpose of this study is to maximize the progression band, the robust deviation decision will be found by modifying Equation 4.33 as follows:

$$z_D = Min\ y \quad 4.34$$

subject to:

$$z^s - f(X, D^s) \leq y \qquad s \in S,\ X \in \cap_{s \in S} F_s \quad 4.35$$

The bandwidth optimization problem given in the previous section is solved first to obtain the optimal solutions for all possible traffic volumes. Next, the robust discrete optimization problem given in Equations 4.34 and 4.35 is solved along with the constraints of the bandwidth optimization problem.

As seen in the review of literature, the objective of progression techniques is not to lower system delays or the number of stops. Instead, progression models try to minimize disruption to arterial traffic flows. The proposed system aims to capitalize on the benefits of progression control methods and optimization strategies by integrating them in a real-time operational system. This is done by using bandwidths generated at the *progression* or *network control level* as constraints to the signal optimization problem, which is solved at the *intersection control level*. The *intersection control level* will be discussed in the next section. The *network control level* will pass the following control variables to the *intersection control level*:

- progression bands for each link and in each direction
- system cycle length

Once the cycle length nears completion, the bandwidth maximization problem is solved again at the *progression control level*, which then determines the control variables and passes the information to the *local level*. A schematic representation of the interrelations between the component modules of the *progression control level* is given in Figure 4.3.

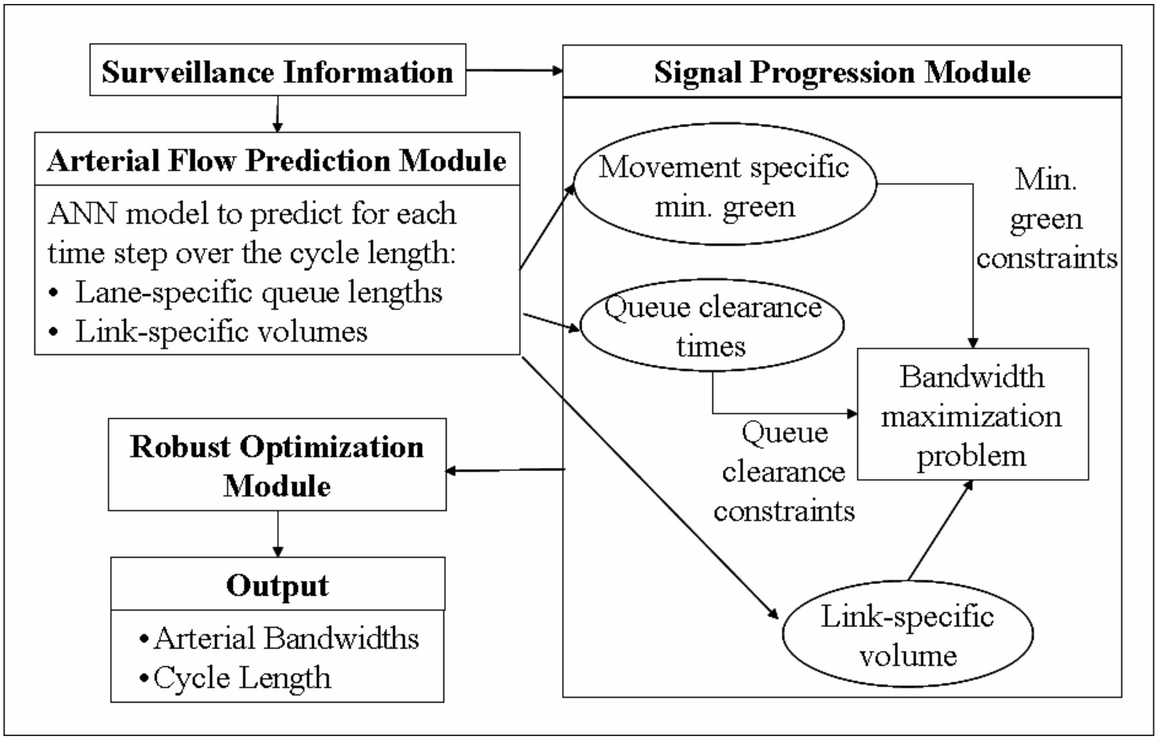

Figure 4.3 Component Modules of the Progression Control Level

4.4 LOCAL OR INTERSECTION CONTROL LEVEL

This section presents the *local* or *intersection control level*. At the *local* or *intersection control level*, the signal operations at each intersection are optimized within the progression bands generated by the *network control level*. This is to ensure that progression is not disrupted along the arterial, while optimizing the phasing plans using a rolling horizon approach. The optimization is done over a time horizon equal to the cycle length of the intersection, computed at the *network control level*. The signals are optimized at an interval of 2 to 5 seconds. The objective of the optimization technique is to minimize a weighted-combination of vehicle queue lengths, delays and stop times, with constraints on the bandwidth and minimum green time. This will ensure that the platoons arriving at an intersection will get the allocated bandwidth generated at the *network control level*, and no movement suffers from excessive delays.

It should be noted that robust discrete optimization technique is not used at this level, since signal optimization is performed only for a single intersection over a short period of time equal to 2 to 5 seconds. Due to the short planning horizon, uncertainty in traffic data is limited and any variations can be captured quickly. Robust optimization provides a robust decision that may not be optimal. Hence, given that the bandwidth solution is robust, the signal timings are adjusted within the bands to optimize the performance function.

To satisfy all the functional requirements stated in Chapter 3, the *intersection control level* is designed to have the following features:

- **Signal timing plans are optimized using a rolling horizon approach.**

The optimization is performed over a time horizon equal to the cycle length of the intersection, computed at the *network control level*, but is implemented for only one control time interval (2 to 5 seconds). The signal timing plans are re-optimized at the end of every control time interval. A neural network prediction model is used to predict link and lane-specific queue lengths and vehicle arrivals for each control time interval over the time horizon. Signal timing plans are optimized for each intersection by minimizing a weighted combination of vehicle queue lengths, delays and stops over a time horizon equal to the cycle length. The optimization technique has the ability to skip phases in the event that there is no vehicle demand and is solved using dynamic programming.

- **A minimum green criterion is imposed for each phase.**

This is applied for driver and pedestrian safety and to discharge the average waiting queue during each phase. It is a function of the queue at the intersection, which is predicted using a neural network model.

- **Signal timings are optimized within the bandwidths generated from the progression control module to ensure traffic progression.**

The phase sequences and timings in this study are determined to maintain the progression bands obtained from the *network control level*. The signal plans will be generated so that the total green time allocated for the through movement along the arterial in the outbound (inbound) direction is greater than or equal to the bandwidth for the outbound (inbound) direction. The optimization technique has the ability to skip phases in the event that there is no vehicle demand.

The *intersection control level* is divided into three modules: (1) the traffic state prediction module, (2) the signal state estimation module, and (3) the signal optimization module. The major function of each module is described below. Table 4.2 lists the notations used in this level.

The signal state estimation module monitors the signal state, the elapsed green time for each phase and the elapsed arterial through green, and estimates the minimum green duration. The signal optimization module handles the optimization of the signal by minimizing vehicle queue lengths, delays and stop times subject to bandwidth and minimum green constraints.

4.4.1 Traffic State Prediction Module

The traffic state prediction module generates the vehicle queues lengths, vehicle arrivals and other traffic flow measures needed for solving the signal optimization problem at the intersection control level. Data from the surveillance system are used as input to the neural network model developed at the progression control level (Section 4.3.2) to predict vehicle delays, stop times and lane specific queue lengths for each approach link at each intersection along the arterial. Every time signal optimization takes place, the traffic measures are predicted for each control interval over the time horizon.

Table 4.2 Notations Used in the Local Control Level

C	:	System cycle length (seconds)
T	:	Total number of discrete control time intervals (steps) in the time horizon
α	:	Duration of a control time interval (seconds)
H	:	Set of signal phases at the control intersection
P^i	:	Set of lane groups in phase i
$\Phi^i(t)$	:	$\begin{cases} 0 & \text{if signal state is green for phase } i \text{ and control time interval } t \\ 1 & \text{if signal state is red for phase } i \text{ and control time interval } t \end{cases}$
$\xi^i(t)$	:	$\begin{cases} 1 & \text{if existing signal state of phase } i \text{ is switched at the end of control time step } t \\ 0 & \text{otherwise} \end{cases}$
$Q_l^{i,k}(t)$	:	Queue length at intersection k in lane l of phase i at time t
t_s^p	:	Start-up lost time for automobiles (seconds)
G_{min}^i	:	Pre-specified minimum green time for phase i (seconds)
$G_{min}^{i,k}$	:	Minimum green for phase i at intersection k (seconds)
$U^i(t)$	:	Green time used by phase i, at the end of control time interval t (seconds)
Y^i	:	Yellow time for phase i (seconds)
AR^i	:	All red time for phase i (seconds)
$S_l^{i,k}$	:	Saturation flow rate for lane l of phase i at intersection k (vphpgpl)
$G^{a,k}(t)$	:	Total green time allocated for outbound through traffic at intersection k at the end of control time interval t (seconds)

$\overline{G}^{a,k}(t)$: Total green time allocated for inbound through traffic at intersection k at the end of control time interval t (seconds)

b_k ($\overline{b}_k$) : Outbound (inbound) bandwidth between intersections k and $k+1$, measured in units of the cycle length, C

$w_{q,l}^{i,k}$: Weight associated with vehicle queue length on lane l of phase i at intersection k

$w_{d,l}^{i,k}$: Weight associated with vehicle delay on lane l of phase i at intersection k

$w_{s,l}^{i,k}$: Weight associated with vehicle stop times on lane l of phase i at intersection k

$d_l^{i,k}(t)$: Vehicle delays at control time interval t on lane l of phase i at intersection k

$s_l^{i,k}(t)$: Vehicle stop times at control time interval t on lane l of phase i at intersection k

s_n : Total number of control time intervals that have been allocated after *stage n* is completed (*state variable*)

$x_n(s_n)$: Number of control time intervals allocated for green time for *stage n* (*decision variable*)

$Q_l^{i,k}(s_n)$: Queue length at intersection k in lane l of phase i during *stage n* (*transformation function*)

$d_l^{i,k}(s_n)$: Vehicle delays at intersection k in lane l of phase i during *stage n* (*transformation function*)

$s_l^{i,k}(s_n)$: Vehicle stop times at intersection k in lane l of phase i during *stage n* (*transformation function*)

$r_n(s_n)$: Cumulative value of the performance index for *stage n* (*return function*)

4.4.2 Signal State Estimation Module

This module serves to monitor the signal state, compute the elapsed green time for each phase and the elapsed arterial through green, and estimate the minimum green duration. Some key functions performed by this module are given below.

Elapsed Green Computation

The green time that will be used by phase *i*, at the end of control time interval *t*, is computed as follows:

$$U^i(t) = (U^i(t-1) + \alpha)(1 - \xi^i(t-1)) \quad \forall i \in H \tag{4. 36}$$

Based on the control decision, $\xi^i(t\text{-}1)$, green time is either increased by α seconds, which is the duration of a control time interval, or is set to zero.

Minimum Green

Minimum green is the shortest green time during which drivers can be expected to react safely to signal changes. In this study it is also defined to be sufficiently long for discharging the average waiting queue during each control phase *i*. A mathematical representation of such a requirement is given below.

$$G_{min}^{i,k} = Max\left\{ \left(t_s^p + \frac{3600}{S_l^{i,k}} \underset{t}{Max}\left(\underset{l}{Avg}\ Q_l^{i,k}(t) \right) \right), G_{min}^i \right\} \tag{4. 37}$$

$$\forall l \in P^i,\ \forall i \in H,\ \forall k$$

Minimum green, $G_{min}^{i,k}$ for phase *i* at intersection *k* is defined to be the maximum of the following two terms:

- start up lost time due to switching of signals plus the time taken to discharge the maximum of the average queue lengths for all lanes in phase *i* at control time interval *t*
- pre-specified minimum green time to allow drivers to react safely to signal changes and for pedestrians to safely cross the street

Note that, the queue lengths used in the computation of minimum green is provided by the Traffic State Prediction Module.

Arterial Green

Arterial green serves as a counter to ensure that the *intersection control level* will provide the coordination bandwidth allocated by the *progression control level* to the platoons traveling along the arterial in both directions. Outbound (inbound) arterial green, $G^{a,k}(t)$ ($\overline{G}^{a,k}(t)$) is the total green time allocated for through traffic in the outbound (inbound) direction along the arterial. This is computed for each intersection *k* for each control time interval *t*. When the control is switched from a through phase in the outbound (inbound) direction along the arterial to side streets or to a non-through movement along the arterial, $G^{a,k}(t)$ ($\overline{G}^{a,k}(t)$) is set to zero. Otherwise, it is increased by α seconds, which is the duration of a control time interval.

4.4.3 Signal Optimization Module

This module functions to optimize the signal timings plans within the bandwidth constraints using dynamic programming.

Signal Optimization

The objective of the optimization problem is to minimize a weighted sum of vehicle queue lengths, delay and stop times. The problem is formulated as follows:

$$Minimize\ PI = \sum_{t=1}^{T} \sum_{i \in H} \sum_{l \in P^i} \left(w_{q,l}^{i,k}\, Q_l^{i,k}(t) + w_{d,l}^{i,k}\, d_l^{i,k}(t) + w_{s,l}^{i,k}\, s_l^{i,k}(t) \right) \qquad 4.38$$

subject to:

$$\xi^i(t)(1-\xi^i(t)) = 0 \qquad \forall i \in H, t \in [1,T] \qquad 4.39$$

$$\Phi^i(t)(1-\Phi^i(t)) = 0 \qquad \forall i \in H, t \in [1,T] \qquad 4.40$$

$$U^i(t) - G^{i,k}_{min} \geq 0 \qquad \forall i \in H, t \in [1,T] \qquad 4.41$$

$$G^{a,k}(t) \geq b_k\ C \qquad t \in [1,T] \qquad 4.42$$

$$\overline{G}^{a,k}(t) \geq \overline{b}_k\ C \qquad t \in [1,T] \qquad 4.43$$

Equations 4.39 and 4.40 define the binary variables $\xi^i(t)$, which is the control decision and $\Phi^i(t)$, which is the signal state. Equation 4.41 is the minimum green constraint. A minimum green criterion is imposed on the green time for each phase and for each control time interval. This will ensure that side streets do not suffer when arterial volumes are high, since minimum green time is computed to be the time required for discharging the average queue.

The control decision, $\xi^i(t)$, the signal state, $\Phi^i(t)$, the elapsed green time, $U^i(t)$, and the minimum green, $G^{i,k}_{min}$ are provided by the Signal State Estimation Module. Equations 4.42 and 4.43 are the bandwidth constraints. The bandwidth constraints ensure that the total green time allocated for through movement is greater than the bandwidth computed by the *progression control level*. The green time allocated for the through movement is provided by the Signal State Estimation Module, and the bandwidths are supplied by the *progression control level*.

Solution Algorithm

The optimal signal timing plan is computed using dynamic programming approach, since it is based on decomposing a complex problem into a series of smaller problems, and solving them systematically to determine the best solution that maximizes the overall benefit. The advantage of dynamic programming approach over sequential search methods is that it determines absolute or global optima whereas the sequential search method can only guarantee relative or local optima. This is because the dynamic programming approach is based on the Bellman's (1957) "principle of optimality" which can be stated as:

> "*An optimal policy has the property that whatever the initial state and decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision*".

In this study, the optimization horizon, which is equivalent to the system cycle length C, is divided into T control time intervals each of duration α seconds. The signal optimization problem is solved using dynamic programming approach by making the following interpretations:

- Each signal phase corresponds to a *stage*.
- The total number of control time intervals that have been allocated at the end of *stage n* corresponds to the *state variable*, s_n.
- The green time allocated to each *stage n* corresponds to the *decision variable*, $x_n(s_n)$. Note that, $x_n(s_n)$ is expressed in terms of the number of control time intervals that are allocated for the green time.
- The measures of effectiveness (i.e., queue lengths, stop times and delays) correspond to the *transformation function*.

- The cumulative value of the performance indices (i.e., the weighted sum of the queue lengths, delays and stops) corresponds to the *return function*.

<u>State variable</u>

The *state variable* or the total number of control time intervals that have been allocated at the end of *stage n*, s_n, is defined to be the sum of the following four terms:

- number of control time intervals that have been allocated in the previous stage, i.e., *n-1*
- number of control time intervals that have been allocated for the green time in *stage n*
- number of control time intervals that have been allocated for the yellow time in *stage n*
- number of control time intervals that have been allocated for the clearance time in *stage n*

A mathematical representation of the above is as follows:

$$s_n = s_{n-1} + x_n(s_n) + \frac{Y^i}{\alpha} + \frac{AR^i}{\alpha} \tag{4. 44}$$

By including the yellow time and clearance interval in the above equation, if a decision is made during signal optimization to terminate the current phase, a minimum of the sum of the yellow time and the clearance interval will be guaranteed.

Decision variable

Let $X(x_n)$ be the set of feasible values of the decision variable, $x_n(s_n)$. Then we can define $X(x_n)$ as follows:

$$X(x_n) = \left\{0, \frac{G_{min}^{i,k}}{\alpha}, \left(\frac{G_{min}^{i,k}}{\alpha} + 1\right), \left(\frac{G_{min}^{i,k}}{\alpha} + 2\right), \ldots, \left(s_n - s_{n-1} - \frac{Y^i}{\alpha} - \frac{AR^i}{\alpha}\right)\right\} \quad 4.45$$

By allowing $x_n(s_n)$ to assume a value of '0', it is possible to skip a phase in the event that there is no demand. If $x_n(s_n)$ is non-zero, then the minimum number of control time intervals that will be allocated is equal to $\frac{G_{min}^{i,k}}{\alpha}$, thereby satisfying the minimum green constraint given in Equation 4.45.

Transformation functions

The following three transformation functions are defined in the optimization problem - (1) vehicle queue lengths, (2) vehicle delays, and (3) vehicle stop times.

(1) Vehicle queue lengths: Queue lengths for each phase are obtained from the ANN prediction model. To solve the optimization problem, queue lengths are estimated for each feasible value of $x_n(s_n)$.

(2) Vehicle delays: There are numerous components of vehicle delays, such as stopped delay, intersection control delay, approach delay, time-in-queue delay, travel-time delay. In the 1997 version of HCM, the intersection control delay was defined as follows: "Control delay includes initial deceleration delay, queue move-up time, stopped delay and final acceleration delay". Intersection control delay is caused by the intersection signal control. Travel-time delay is the difference between the desired travel time and the actual travel time. It is

affected by the geometry, signal control, parking activities, bus operations, etc. Stopped delay only considers the time lost while a vehicle is stopped in queue waiting for a green signal indication or waiting for its leader to move forward. It does not consider the time lost while the vehicle is slowing down, and approaching the stop bar or the end of the queue, nor does it consider the time lost while the vehicle accelerates to its normal operating speed. Approach delay accounts for the time lost in decelerating, the time that the vehicle is stopped in queue, and the time spent in accelerating, but the delay accumulation stops, when the vehicle clears the stop bar. It does not consider the time lost to attain its normal operating speed beyond the stop bar. The time-in-queue delay is the time spent by the vehicle in queue, and like the stopped delay, it does not account for the time spent decelerating to join the queue, nor the time spent accelerating to its normal speed. Hence, since travel-time delay is not dependent only on the signal control and since stopped delay, in-queue delay, and approach delay under-estimate the total delay caused by the signal control, intersection control delay is used to represent the vehicle delay.

(3) Vehicle stop time: Vehicle stop time is the amount of time a vehicle is stopped on the link.

The optimal phase durations and sequences are determined by solving the signal optimization problem using the following solution algorithm:

Step 1: Set $n = 1$. Go to next step.

Step 2: Determine all feasible values for $x_n(s_n)$ using Equation 4.45. Go to next step.

Step 3: For each feasible value, determine the transformation functions,

$Q_l^{i,k}(s_n)$, $d_l^{i,k}(s_n)$, and $s_l^{i,k}(s_n)$. Go to next step.

Step 4: Compute the *return function* as follows:

$$r_n(s_n) = \sum_{t=1}^{s_n} \sum_{i \in H} \sum_{l \in P^i} \left(w_{q,l}^{i,k}\, Q_l^{i,k}(t) + w_{d,l}^{i,k}\, d_l^{i,k}(t) + w_{s,l}^{i,k}\, s_l^{i,k}(t) \right)$$

Go to next step.

Step 5: Determine the optimal value for the decision variable $x_n^*(s_n)$ that minimizes the return function given in Step 4. Go to next step.

Step 6: Set $s_n = s_n + 1$. If $s_n \leq T$, go to Step 2. Otherwise, go to next step.

Step 7: If $n \leq |H|$, set $n = n+1$ and $s_n = \dfrac{Y^i + AR^i}{\alpha}$. Go to Step 2.

Otherwise, go to next step.

Step 8: Determine the optimal phase timings and sequence that minimize the *return function*. Go to next step.

Step 9: Check if the bandwidth constraints are satisfied:

$$G^{a,k}(t) \geq b_k \qquad t \in [1,T], \forall k$$

$$\overline{G}^{a,k}(t) \geq \overline{b}_k \qquad t \in [1,T], \forall k$$

If the constraints are satisfied, select the phase timings and sequence determined in Step 8 as the optimal solution. Otherwise, go to Step 8 and determine the next best solution.

Thus, the arterial signal problem is solved by maximizing the bandwidths once every cycle length, and optimizing the signal timing plans at each individual intersection once every control time interval. Figure 4.4 presents the interrelations between the component modules of the *intersection control level*. When the cycle length is nearing completion, the bandwidths are maximized again and the future optimum cycle length is determined. While the *progression control level* optimizes the through bands, the *local level* continues to optimize the signals till the end of the current cycle length. The system treats a bus as an automobile unless detected by the bus surveillance system. Once a bus is detected, the control is passed to the *bus priority control level*, which is presented in the next chapter. The *intersection control level* passes the optimal signal timing plans at each intersection to the *bus priority control level*.

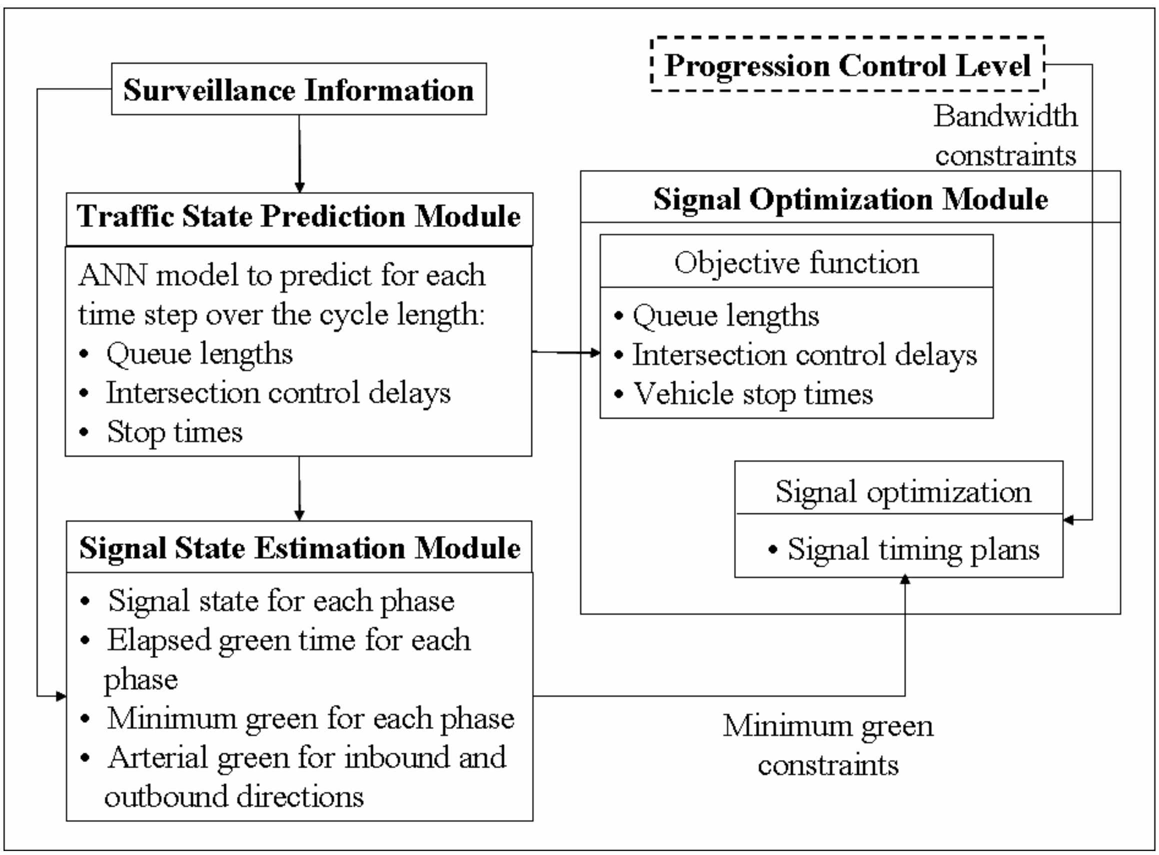

Figure 4.4 Component Modules of the Intersection Control Level

4.5 SUMMARY

In this chapter, a robust optimization model for arterial signal control was presented. The proposed system was decomposed into two levels, the *network* or *progression control level* and the *local* or *intersection control level*.

The *progression control level* served to provide progression along the arterial. The arterial progression model was a modified version of the MULTIBAND bandwidth maximization model. Unlike other bandwidth maximization models, the proposed model did not use a pre-specified queue clearance time; instead, the queue clearance time was computed as a function of the existing queue at the intersections. An artificial neural network prediction model was used to predict the vehicle arrivals and queues at each intersection along the arterial. The model imposed a minimum green criterion for each phase, thereby ensuring that side streets would not get penalized and safety of pedestrians would not be compromised when arterial volumes were high.

Most of the existing arterial progression strategies do not use real-time traffic data and hence are deficient in real-time applications. The proposed model eliminated this drawback by using real-time surveillance information for training the artificial neural network queue prediction model. However, the model performed the bandwidth optimization only once during every system cycle length. To eliminate uncertainty in the prediction due to the long time horizon and large system, robust discrete optimization technique was employed to determine the progression bands. Robust discrete optimization guaranteed that the generated progression bands were the best over all realizable traffic scenarios for the given time horizon.

In the *local* or *intersection control level*, the signal timing plans and phase sequences were optimized for each intersection, based on the coordination bandwidth constraints. Minimum green criterion was imposed for each phase for driver and pedestrian safety. The optimal signal timings were obtained using dynamic programming method by minimizing a performance index, which was a weighted sum of vehicle stop times, queue lengths, and intersection control delays.

Note that robust optimization technique was not used at this level since the time horizon was very short. If reliable surveillance information is present, traffic data can be suitably predicted with very little uncertainty for a control period of 2 to 5 seconds and if any variations exist, the optimization model can capture them in the next control period. Bus priority under arterial progression will be presented in the next chapter.

CHAPTER 5: A REAL-TIME BUS PRIORITY CONTROL UNDER SIGNAL PROGRESSION

5.1 INTRODUCTION

The objective of this chapter is to present a bus priority control model under signal progression control, presented in the previous chapter. In the proposed system, unconditional priority is not given to buses over passenger cars. Unlike most conventional transit priority systems which give priority to buses without considering the effect on other traffic users, the priority decision is based on a performance index which includes automobile and bus passenger delays, vehicle delays and bus schedule delays.

The models are developed under incident free conditions. If an incident occurs, then bus priority is no longer justifiable. The problem is entirely different until congestion is cleared and traffic flow has been stabilized.

In this chapter, the key features and a detailed description of the bus priority control model that is used in the *bus priority control level* will be presented in Section 5.2. The solution algorithm will be discussed in Section 5.3 and the chapter will be summarized in Section 5.4.

5.2 BUS PRIORITY CONTROL MODEL

In this section, the bus priority control model is presented. The objective of the *bus priority control level* is to provide priority to buses under arterial signal progression, without disregarding other traffic. To satisfy all the functional requirements stated in Chapter 3, the bus priority control model is designed to have the following key features:

- **Bus priority is incorporated as one of the adaptive signal control functions.**

The decision to grant priority by modifying signal timing plans is made by minimizing a performance index, which is a weighted sum of vehicle delays, the bus schedule delay, and delays incurred by automobile and bus passengers. A minimum green constraint is imposed, whether there is a bus demand or not, for all approaches, thereby minimizing unfair treatment of automobiles in the presence of a bus. It is defined as a function of the vehicle queue length which is predicted using a neural network prediction model.

- **Bus schedule adherence is included in the priority decision.**

This is achieved by minimizing a performance index, which is function of bus schedule delays.

- **Bus priority is provided without discounting signal progression.**

The decision to give priority treatment to buses is made based on performance index subject to bandwidth constraints. By including bandwidth constraints, determined at the progression control level, the system is able to afford bus priority by considering its effect on arterial progression.

In the proposed bus priority control system, real-time arrival information is assumed to be available from existing surveillance systems, such as AVL/CAD systems. If any bus is identified by the surveillance system, the bus priority control model evaluates whether the bus warrants any priority, and the decision is made using a performance index. The notations used in the study are presented in Table 5.1. The *bus priority control level* is divided into three component modules: (1) bus prediction module, (2) priority candidacy module, and (3) bus priority optimization module. The major function of each module is described below in detail.

Table 5.1 Notations Used in the Bus Priority Control Level

C	:	System cycle length (seconds)
T	:	Total number of discrete control time intervals (steps) in the time horizon
α	:	Duration of a control time interval (seconds)
C^{i}_{pd}	:	Passenger delay for phase *i*
C^{i}_{vd}	:	Vehicle delay for phase *i*
C^{i}_{bd}	:	Bus schedule delay for phase *i*
c_p	:	Unit cost or weight associated with passenger delay
c_v	:	Unit cost or weight associated with vehicle delay
c_b	:	Unit cost or weight associated with bus schedule delay
H	:	Set of signal phases at the control intersection
P^i	:	Set of lane groups in phase *i*
s_n	:	Total number of control time intervals that have been allocated after *stage n* is completed (*state variable*)
$BQ^{i,k}_{l}(s_n)$	:	Number of buses competing for priority on lane *l* of phase *i* at intersection *k* at the end of *stage n* (vehicles)
$p^{i,k}_{l,j}$	:	Number of passengers on bus *j* on lane *l* of phase *i* at intersection *k*
p_a	:	Average occupancy of automobiles
$PQ^{i,k}_{l}(s_n)$	:	Total number of automobiles in queue on lane *l* of phase *i* at intersection *k* at the end of *stage n* (vehicles)
$SD^{i,k}_{l,j}(t)$	:	Schedule delay of bus *j* on lane *l*, at time step *t* for phase *i* of intersection *k* (seconds)
$\Phi^{i}(t)$	:	$\begin{cases} 0 & \text{if signal state is green for phase } i \text{ and control time interval } t \\ 1 & \text{if signal state is red for phase } i \text{ and control time interval } t \end{cases}$
$\xi^{i}(t)$	:	$\begin{cases} 1 & \text{if existing signal state of phase } i \text{ is switched at the end of control time step } t \\ 0 & \text{otherwise} \end{cases}$
$G^{i,k}_{min}$	:	Minimum green for phase *i* at intersection *k* (seconds)
$U^{i}(t)$	:	Green time used by phase *i*, at the end of control time interval *t* (seconds)

$G^{a,k}(t)$	:	Total green time allocated for outbound through traffic at intersection *k* at the end of control time interval *t* (seconds)
$\overline{G}^{a,k}(t)$	:	Total green time allocated for inbound through traffic at intersection *k* at the end of control time interval *t* (seconds)
b_k ($\overline{b}_k$)	:	Outbound (inbound) bandwidth between intersections *k* and *k+1*, measured in units of the cycle length, *C*

5.2.1 Bus Prediction Module

In this research, a bus warrants priority only when it clears the bus stop. For each bus that has cleared the bus stop, it is necessary to estimate the time when it will join the current queue or reach the stop bar. This is done through a neural network model which is developed for predicting all bus-related data that are pertinent to the bus priority control model. Real-time surveillance information is used to train the neural network model. As AVL information was not available for conducting this study, CORSIM is used to generate data for training the ANN model. The following bus-related information is used as input for training the neural network:

- Bus location
- Bus speed
- Time when bus has finished dwelling at the bus stop

Note that information such as bus location, speed and schedule delay are available from the AVL/CAD system, and hence are known at any given instance. However, since the bus priority model is optimized every control time interval over the entire time horizon, it becomes necessary to predict the bus speed and estimate the bus location for subsequent control time intervals within the time horizon. Figure 5.1 shows the structure of the ANN model.

Once a bus is detected by the surveillance system, the control is passed from the *local level* to the *bus priority level*. All predicted information is passed to the next module, the priority candidacy module.

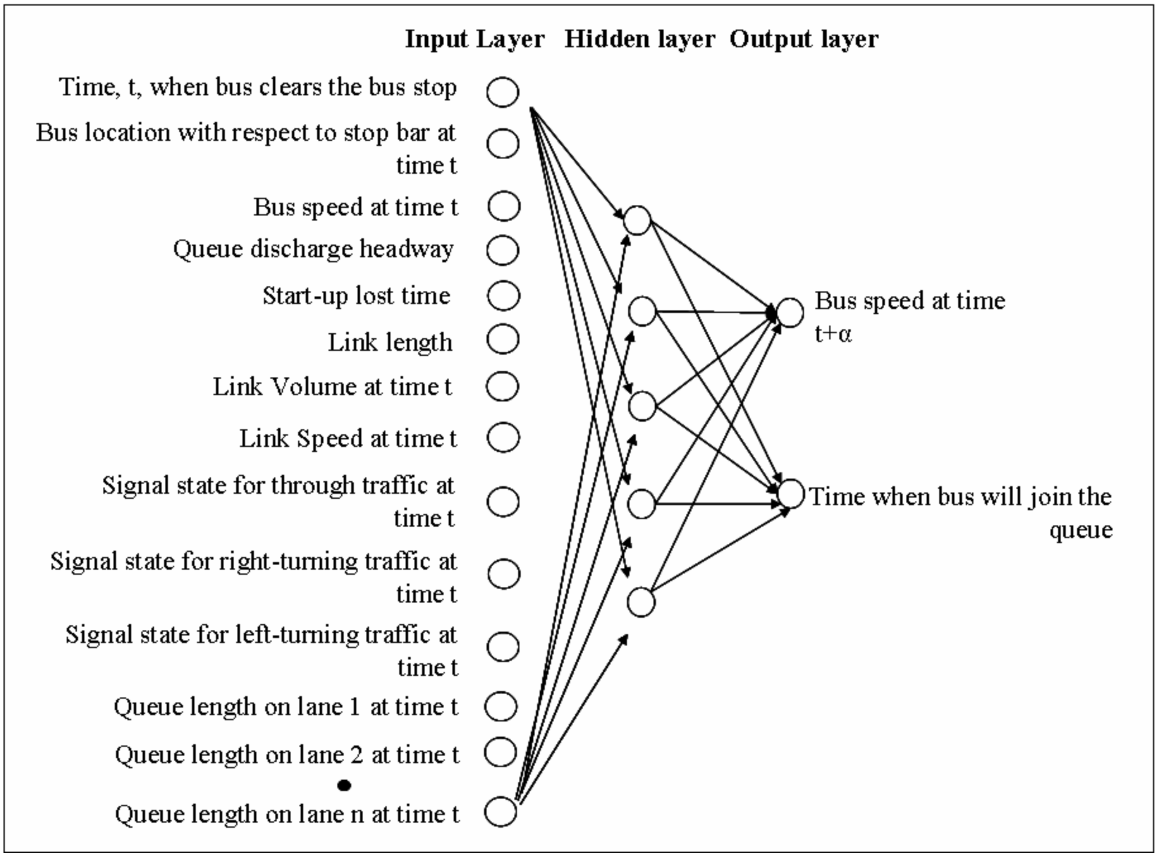

Figure 5.1 Structure of Bus Prediction ANN Model

5.2.2 Priority Candidacy Module

This module's primary function is to check if the detected bus is eligible to compete for priority treatment based on the information provided by the bus prediction module. The decision is made based on the following rules:

1. Check if the bus needs to stop at a downstream bus stop on the link.

   If it is true, then the bus cannot compete for priority treatment.

   Otherwise, the bus was either detected after it cleared the bus stop, or it does not need to stop at the bus stop. Check the next rule.

2. Check if the bus is heading towards the bus depot after completing its routes.

   If it is true, no special priority treatment need be considered.

   Otherwise, it may warrant a priority treatment. The bus can compete for priority.

If the bus does not require any special priority treatment, the existing signal timing plan generated at the *local level* will continue to be implemented as the optimal signal timing plan. If a decision is made that the detected bus can compete for priority, the control is passed to the next module, the bus priority optimization module, to evaluate whether the existing signal-timing plan needs to be modified.

5.2.3 Bus Priority Optimization Module

It has been seen that most adaptive control strategies do not consider the schedule delay of a bus while making a signal state decision for bus priority. Hence, the decision to switch over to another phase, may not be an optimal solution. This can be rectified by computing a performance index (PI) that evaluates the effect of the decision. Since the objective of this study is to develop a bus priority control model for arterial streets, the performance index will also consider the effect of priority treatment on signal progression. To satisfy these two requirements, a PI model, allowing for measuring the benefit of the control decision, based on bus schedule delays, vehicle delays and passenger delays, subject to arterial bandwidth constraints is formulated in this section.

The performance index is expressed as the weighted sum of automobile and bus passenger delays, C_{pd}^{i}, vehicle delays, C_{vd}^{i}, and bus schedule delays, C_{bd}^{i}, where the weights attached can be chosen to be the unit time costs associated with each category of delay. Obtaining a reliable unit time cost for the three categories of delays is beyond the scope of the research, but the proposed model allows users to assign their own values to the unit time costs. The weights can also be representative of the prevailing traffic conditions. The performance index can thus be written as:

$$PI = \sum_{i \in H} \left(c_p \, C_{pd}^{i} + c_v \, C_{vd}^{i} + c_b \, C_{bd}^{i} \right) \qquad 5.\,1$$

Computation of Passenger Delay

Passenger delay consists of delay caused to passengers of automobiles and buses. By including passenger delays, it is possible to afford each bus passenger equal treatment as a passenger in an automobile. It can be expressed as follows:

$$C_{pd}^{i} = \sum_{l \in P^i} \sum_{j=1}^{BQ_l^{i,k}(s_n)} p_{l,j}^{i,k} + p_a \sum_{l \in P^i} PQ_l^{i,k}(s_n) \qquad 5.\,2$$

The first term is the number of bus passengers. This is computed as the sum of the occupancy of all buses that are in queue. The second term represents the total occupancy of all automobiles that are in queue. It is computed by multiplying the average occupancy of automobiles, p_a with the total number of automobiles in queue.

Computation of Vehicle Delay

As discussed previously in Section 4.4.2, intersection control delay results from the intersection signal control. Hence, at the bus priority control level also intersection control delay is used to characterize vehicle delays. This is obtained directly from the output generated by CORSIM.

Computation of Schedule Delay

Bus schedule delay is computed as the sum of the schedule delays of all buses that are unable to discharge during *stage n*. The bus schedule delay for each individual bus is computed as the sum of the schedule delay that the bus is currently experiencing, and any additional delay that it experiences due to the signal control. It can be written as:

$$C_{sd}^{i} = \sum_{j=1}^{BQ_{l}^{i,k}(s_n)} \left(SD_{l,j}^{i,k}(t) + \alpha \left(s_n - t \right) \right) \qquad 5.\ 3$$

Buses that are unable to discharge either due to a red phase or due to queued vehicles in front of it, have a schedule delay that is equal to the sum of the schedule delay that it is currently experiencing at time step *t* and the additional time from time *t* to the end of *stage n*. It is likely that a bus may arrive on time or is ahead of its schedule. In such situations, a bus does not warrant a priority. Hence, the schedule delay is taken as zero if the bus is on time and negative if it is ahead of schedule.

Signal Optimization Problem

The objective of the optimization problem is to minimize the performance index, PI, which is a weighted sum of passenger delays, C_{pd}^{i}, vehicle delays, C_{vd}^{i}, and bus schedule delays, C_{bd}^{i}. The problem is formulated as follows:

$$Minimize\ PI = \sum_{i \in H} \left(c_p\, C_{pd}^{i} + c_v\, C_{vd}^{i} + c_b\, C_{bd}^{i} \right) \qquad 5.\ 4$$

subject to:

$$\xi^i(t)(1-\xi^i(t)) = 0 \qquad \forall i \in H, t \in [1,T] \qquad 5.5$$

$$\Phi^i(t)(1-\Phi^i(t)) = 0 \qquad \forall i \in H, t \in [1,T] \qquad 5.6$$

$$U^i(t) - G^{i,k}_{min} \geq 0 \qquad \forall i \in H, t \in [1,T] \qquad 5.7$$

$$G^{a,k}(t) \geq b_k C \qquad t \in [1,T] \qquad 5.8$$

$$\overline{G}^{a,k}(t) \geq \bar{b}_k C \qquad t \in [1,T] \qquad 5.9$$

Equations 5.5 and 5.6 define the binary variables $\xi^i(t)$, which is the control decision and $\Phi^i(t)$, which is the signal state. Equation 5.7 is the minimum green constraint. The minimum green constraint is imposed on the green time for each phase and for each time step. Equations 5.8 and 5.9 are the bandwidth constraints to ensure that bus priority does not disrupt signal progression. A detailed presentation of the computation of the elapsed green, minimum green and arterial bandwidths are given in Section 4.4.2 of the previous chapter.

Solution Algorithm

The bus priority model is solved by minimizing the performance index given in the previous section using dynamic programming approach, which was discussed in detail in Section 4.4.3. The transformation functions for the bus priority control model are the passenger delays, control delays, and bus schedule delays.

Figure 5.2 illustrates the interrelations between the component modules of the *bus priority control level*.

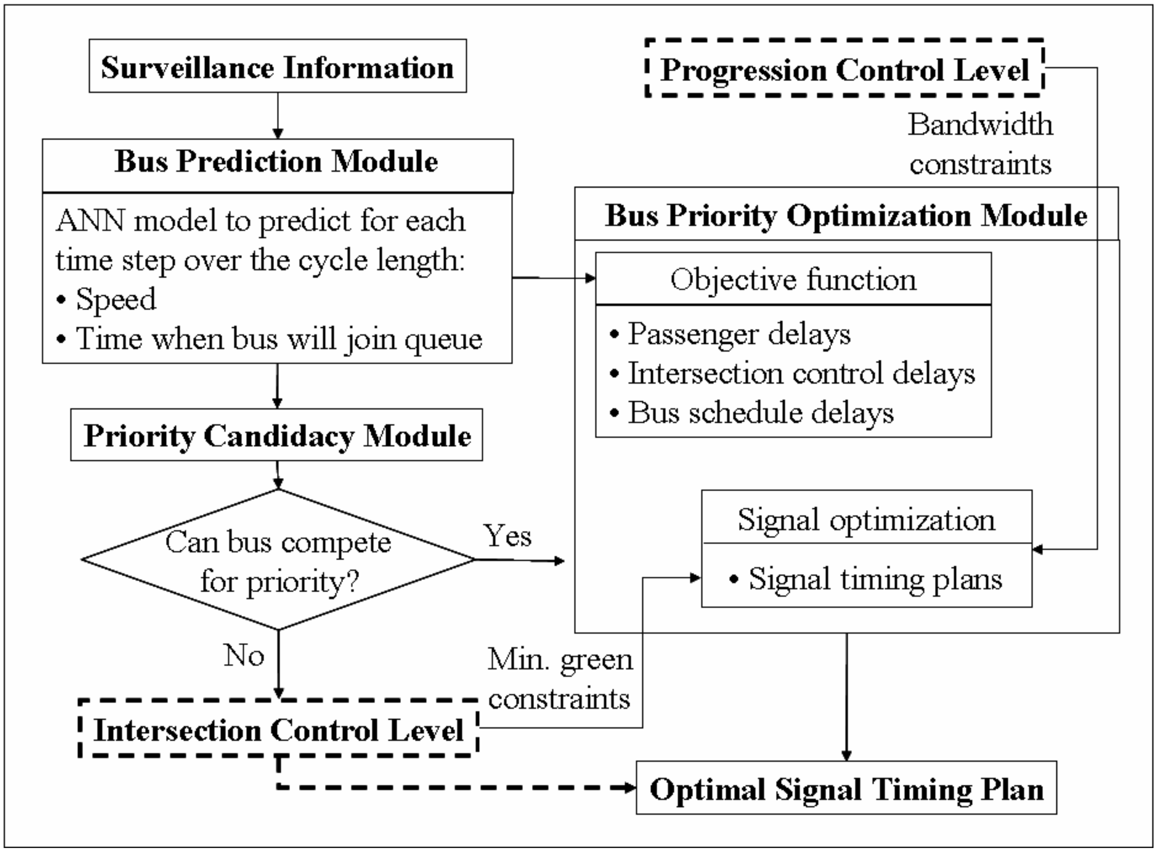

Figure 5.2 Component Modules of the Bus Priority Control Level

5.3 SUMMARY

In this chapter, the bus priority model was formulated. Unlike most conventional bus priority strategies, the model did not give absolute priority to buses. Driver safety and overall minimization of queue length were two main deciding factors while imposing the minimum green requirement. The control decision for signal setting was based on a performance index, which incorporated schedule delay of a bus, control delay as well as delays incurred by automobile and bus passengers, subject to bandwidth constraints which allowed for uninterrupted arterial progression.

The model had several features that eliminated some of the limitations that were observed in the literature. Very few bus priority systems have been implemented on arterial streets with signal coordination. None of the priority systems dealt with situations where there was more than one bus competing for priority on conflicting approaches. Most systems preempted a bus without evaluating its schedule delay. A bus that is ahead of schedule should be given less priority than an automobile. The proposed systems were able to handle such situations by considering bus schedule delay in the computation of a performance index, which was used in the signal control decision. A bus ahead of schedule was given a negative weight while computing its bus delay. The performance of the bus priority system under robust signal progression is evaluated in the next chapter through a case study.

CHAPTER 6: SYSTEM EVALUATION

6.1 INTRODUCTION

The purpose of this chapter is to evaluate the performance of the integrated system described in Chapters 4 and 5, through a case study. The case study will be conducted in a simulation environment that uses the traffic simulation tool, CORSIM. Two neural network models will be developed for predicting: (i) arterial flows at the *progression control level*, described in Section 4.3.2, and traffic state parameters at the *intersection control level*, described previously in Section 4.4.1, and (ii) bus-related information at the *bus priority control level*, described in Section 5.2.1.

The CORSIM simulation test bed and the experimental design will be described in Section 6.2. The development and testing of artificial neural network models for arterial progression, local intersection optimization, and bus priority operations will be described in Section 6.3. This section will also enumerate the traffic parameters used to train the neural network models. Section 6.4 will discuss the development of a run-time signal controller interface with CORSIM. In Section 6.5, the incremental benefits of the models developed at the three control levels will be evaluated by comparing six signal control alternatives, along with a sensitivity analysis. Finally, the chapter will be summarized in Section 6.6.

6.2 DESIGN OF EXPERIMENTS

This section describes the simulation network and the traffic scenarios that were simulated using CORSIM. CORSIM is a microscopic traffic simulation model that combines two separate programs, one for modeling surface streets and one for modeling freeways. It is a stochastic simulation model that incorporates random processes to model complex driver, vehicle, and traffic system behaviors and interactions. CORSIM is capable of modeling individual vehicle movements on a second-by-second basis for the purpose of assessing the traffic performance of highway and street systems.

CORSIM produces numerous measures of effectiveness, including bus specific output, lane-specific and link-specific outputs, such as volume, control delays, queue lengths, speed, etc. As the output of a stochastic model is random, each run of a stochastic simulation model produces only estimates of a model's true characteristics for a particular set of input parameters. Results from a single run of CORSIM may be misleading. To produce meaningful MOE, several independent runs using various random number seeds are critical. The CORSIM simulation software also permits users to interface with it through an Application Programmers Interface (API), also know as the CORSIM run-time extension. The run-time extension capability allows for the test of adaptive signal control algorithms prior to field deployment. This capability was essential in this study to model the traffic signal control algorithms described in Chapters 4 and 5.

6.2.1 Simulation Network

The simulation network used in this case study is the Speedway Boulevard in Tucson, Arizona. The arterial modeled in this study was 2.67 miles long and had four intersections. To study the impact of the proposed signal progression algorithm on adjacent streets and intersections, an additional 17 intersections that were controlled by actuated signal controllers, were also modeled in the case study. The network also included 24 bus stations, protected as well as unprotected, serviced by eight bus routes that traversed arterial as well as side streets. Figure 6.1 shows the simulation network.

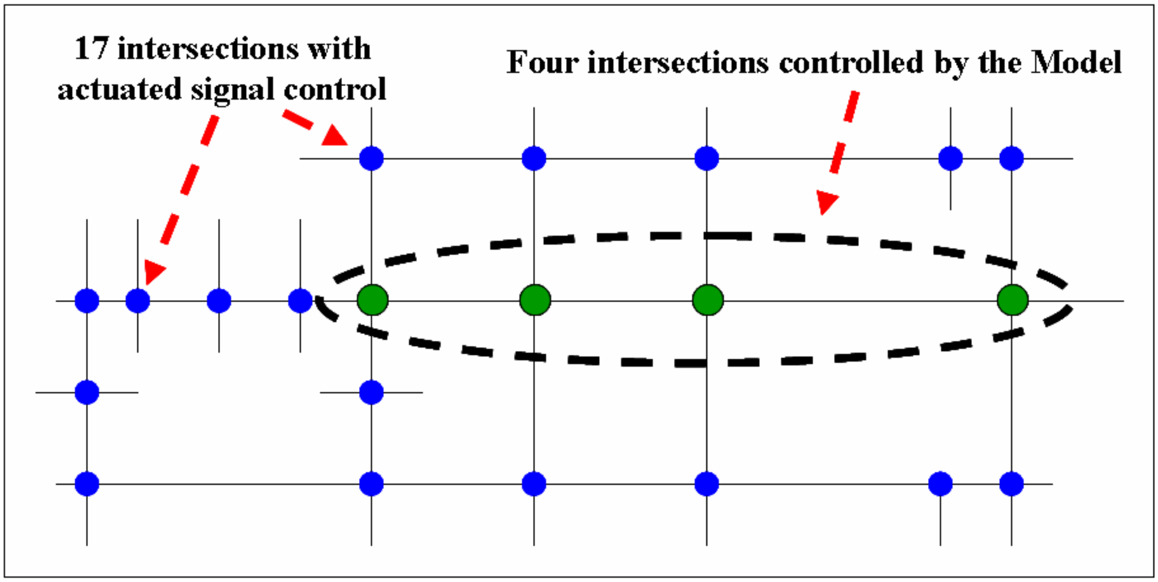

Figure 6.1 The Simulation Network

6.2.2 Bus Data

Although the proposed integrated system relies on information from AVL systems, data such as bus occupancy and bus schedule delays were not available for conducting this research. As this study is experimental in nature, the required data were generated by assuming them to follow a probabilistic distribution. Bus occupancy was assumed to follow a normal distribution with a mean of 30 and a standard deviation of five. Bus occupancy was randomly selected whenever a bus made a stop at one of the 24 bus stations. Bus schedule delays were also assumed to be normally distributed with a mean of five minutes and standard deviation of five. Bus schedule delay was randomly assigned to a bus upon first entering the network.

6.2.3 Surveillance Systems

The operation of the proposed integrated system required the presence of two types of detectors.

Vehicle and Bus Arrival Detectors

These were placed on each lane at a location of 100 ft to 200 ft from the upstream end of each link that was controlled by the proposed integrated system. Vehicle and bus arrivals thus detected were used as input to the neural network models that predicted queue lengths, control delays and stop times. Note that if information from AVL systems is available, bus arrival detectors are not necessary.

Bus Station Clearance Detector

Detectors were placed on each lane 20 ft downstream of the last bus station on each link. A bus would become a candidate for priority consideration when detected by the bus station clearance detector. As noted previously, in the presence of AVL systems, such a detector is not required.

6.2.4 Experimental Scenarios

To evaluate the incremental benefits of the models developed at the three control levels, the following six signal control alternatives were applied to the case study:

(i) Baseline Model: This alternative, modeled the actuated signal control implemented in the field.

(ii) MULTIBAND+: This alternative was modeled to emulate existing arterial signal control models that integrate bandwidth maximization approaches with disutility-based approaches. In this alternative, progression control was provided by the MULTIBAND method, which used pre-defined queue clearance times and static flows. Intersection control was provided by optimizing over one control time interval a performance index, which was a function of vehicle queues, control delays and stop times, subject to bandwidth and minimum green constraints. Bandwidths were provided by the MULTIBAND method. Minimum green times were pre-specified.

(iii) Model I: This alternative was modeled to study the impact of the signal optimization method described in Chapter 4 for providing intersection control. As done in the previous alternative, progression control was provided by the MULTIBAND method. Intersection control was provided by the signal optimization model described in Chapter 4. Optimization was performed over the entire time horizon subject to bandwidth and minimum green constraints. Bandwidths were determined by the MULTIBAND method, and minimum green times were computed as functions of time-varying queues predicted using neural networks.

(iv) Model II: This alternative helped evaluate the performance of the proposed bandwidth maximization approach without robust optimization. The arterial signal control alternative used the bandwidth-maximization model described in Chapter 4 for progression control. This alternative did not make use of robust optimization technique. Intersection control was provided by the signal optimization model described in Chapter 4.

(v) Model III: This alternative was modeled to study the benefits of robust optimization. The arterial signal control alternative used the proposed bandwidth-maximization approach with robust optimization for progression control. Intersection control was provided by the signal optimization model described in Chapter 4.

(vi) Model IV: In this alternative, the robust arterial signal control model described in Chapter 4 was integrated with the bus priority model described in Chapter 5 to evaluate the benefits of bus priority.

The actuated signal alternative employed the existing signal control in the field for all 21 intersections. For the remaining five alternatives, only four out of the 21 intersections were controlled by the proposed signal control models, while the remaining 17 had actuated signal control.

Table 6.1 summarizes the traffic characteristics that were modeled in the case study. The simulation was executed for 1 hour. The control time interval was taken to be five seconds. Note that control time interval is different from the simulation time step, which is specified as one second. Control time interval decides how often the signal optimization takes place at the intersection control level and bus priority control level. For each test case, 10 replications were generated to produce meaningful performance measures, and the statistics averaged over all 10 replications.

Table 6. 1 Summary of Traffic Characteristics Modeled in the Case Study

<table>
<tr><td rowspan="2">Run Control Parameters</td><td colspan="2">Simulation duration</td><td>3600 seconds</td></tr>
<tr><td colspan="2">Simulation time step</td><td>1 second</td></tr>
<tr><td rowspan="6">Traffic Signal Control</td><td colspan="2">Existing Actuated Signal Control</td><td>17 intersections</td></tr>
<tr><td rowspan="5">Proposed signal control alternatives at 4 intersections</td><td>Number of phases</td><td>4</td></tr>
<tr><td>Pre-specified minimum green</td><td>20 seconds</td></tr>
<tr><td>Control time interval</td><td>5 seconds</td></tr>
<tr><td>Yellow</td><td>3 seconds</td></tr>
<tr><td>All red</td><td>2 seconds</td></tr>
<tr><td rowspan="2">Surveillance Detectors</td><td colspan="2">Vehicle and bus arrival detectors</td><td>100 ft ~ 200 ft from upstream end of link</td></tr>
<tr><td colspan="2">Bus station clearance detector</td><td>20 ft downstream of last bus station on link</td></tr>
<tr><td rowspan="4">Traffic Demand Data</td><td colspan="2">EB Arterial link volumes in 15 minute intervals (vph)</td><td>856, 1104, 1672, 1820</td></tr>
<tr><td colspan="2">WB Arterial link volumes in 15 minute intervals (vph)</td><td>1348, 1988, 2208, 2276</td></tr>
<tr><td colspan="2">SB Side street link volumes (vph)</td><td>112 to 1656</td></tr>
<tr><td colspan="2">NB Side street link volumes (vph)</td><td>380 to 1728</td></tr>
<tr><td rowspan="14">Bus Data</td><td rowspan="8">Bus Discharge Headways(s)</td><td>Route 1</td><td>600</td></tr>
<tr><td>Route 2</td><td>600</td></tr>
<tr><td>Route 3</td><td>1800</td></tr>
<tr><td>Route 4</td><td>3000</td></tr>
<tr><td>Route 5</td><td>1800</td></tr>
<tr><td>Route 6</td><td>1800</td></tr>
<tr><td>Route 7</td><td>1800</td></tr>
<tr><td>Route 8</td><td>1800</td></tr>
<tr><td rowspan="4">Bus Station Data</td><td>Number of protected bus stations</td><td>9</td></tr>
<tr><td>Number of unprotected bus stations</td><td>15</td></tr>
<tr><td>Mean dwell time at each bus station</td><td>60 seconds</td></tr>
<tr><td>Bus station bypass percentage</td><td>30%</td></tr>
<tr><td colspan="2">Bus Occupancy</td><td>Normally distributed with $\mu = 30$, $\sigma = 5$</td></tr>
<tr><td colspan="2">Bus Schedule Delay (minutes)</td><td>Normally distributed with $\mu = 5$, $\sigma = 5$</td></tr>
</table>

6.2.5 Measures of Effectiveness

To evaluate the impact of bus priority under signal progression, the following performance measures that were generated by CORSIM were used in the study:

- Control delay (seconds/vehicle): This is defined as the delay incurred by the facility due to the signal control and is the sum of the initial deceleration delay, the queue move-up time, the stopped delay and the final acceleration delay. Average control delays were computed for arterial links as well as side street links.
- Stop times (seconds/vehicle): It is defined as the average time that a vehicle was stopped on the link, and is calculated for both arterial as well as side street links.
- Queue lengths (vehicles): It is defined as the average number of vehicles queued in all lanes of the link. Arterial and side street queue lengths were computed by taking the average of the queued vehicles over all arterial and side street links.
- Speed (mph): Average speeds were computed for arterial and side street links.
- Throughout (vehicles): It is defined as the total number of trips exiting the network. Total throughput was computed only for arterial links.
- Arterial bus delay (minutes/bus): This is defined as the average time that a bus was delayed on an arterial link.
- Side street bus delay (minutes/bus): This is defined as the average time that a bus was delayed on a side street link.
- Bus speed (mph): Average bus speeds were computed for arterial and side street links.

6.3 DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK MODELS

This section discusses the development of two artificial neural network models for predicting traffic data that are essential for executing the following three levels of control: arterial progression, intersection signal optimization, and bus priority operations. The first neural network model (ANN I) was developed to predict for the next control time interval (i.e., looking forward over a period of five seconds) the arterial and side street traffic state variables, including vehicle queue lengths, control delays, stop times, vehicle and bus arrivals, link volumes and link speeds. The second neural network model (ANN II) was developed to predict bus specific data, including bus speeds for the next control time interval and the time when a detected bus would join the queue. Note that the predicted bus speeds were used in estimating the location of the bus at the next control time interval.

Neural networks are characterized by their architecture, the learning or training algorithm, and the activation function. In this study, the feed-forward neural network with back propagation of error was used. The activation function that was used was the sigmoid function, which takes the form, $1/(1+e^{-x})$. The learning rate was specified to be 0.001.

When training a network, the network is continually learning to generalize the relationship between the input and output patterns. When the network starts to memorize the input patterns, and recalls perfectly a pattern that it has learned, memorization or overtraining of the network is said to have occurred. In such situations, the network will not be able to predict the output patterns within the desired level of accuracy when presented with input patterns that are different from those used during training. To prevent overtraining of the network, a second data set, called the validation set, was used to ensure that the neural network model has not simply memorized the patterns within the training set. When the training and validation errors are both seen to decline, the network is still in the mode of generalization, but when the training error continues to decline, and the validation error starts to increase, the memorization or overtraining mode starts. Training of the network should be stopped at this point.

Once a model has been trained to detect, within an acceptable range of accuracy, the output pattern when presented with an input, testing is performed to estimate the generalization error by using a third data set called the test set. This is critical to ensure that the neural network model can be effectively applied to detect general input-output relationships for the intended application. This process is similar to the training process, except that the error is not propagated back through the network.

6.3.1 Creation of Training, Validation and Test Data Sets

To create the training, validation and test data sets, the network described in Section 6.2 was simulated using CORSIM for varying levels of traffic and bus volumes to enable the development of neural network models capable of generalization. Arterial volumes were varied from 500 vph to 2000 vph, and side street volumes were varied from 100 vph to 1800 vph. Bus headways were varied from 300 seconds to 1800 seconds.

Traffic data that were identified as potentially essential for developing the neural network models were extracted from the data structures created by CORSIM using its run-time extension capability. The first ten minutes of data were thrown out, and the remaining data were normalized between 0.25 and 0.85. Of the available patterns, 70% of the data were randomly selected for the design (50% for training and 20% for validation) of the models, and the remaining 30% were reserved as test data.

6.3.2 Design and Test of the Neural Network Models

The procedure to identify the best neural network architecture included two steps. The first step was to identify the input data and the optimum number of hidden units that would give the best results, and the second step was to determine the generalization errors. The criterion for selecting the best model structure was the root mean square error (RMS).

ANN Model for Arterial Flow Prediction and Traffic State Prediction Modules

To identify the input data needed to develop a neural network model (ANN I) capable of predicting arterial and side street flows, two input scenarios were examined. The traffic parameters used in the two scenarios are listed in Table 6.2. Both scenarios used as input, link specific data such as link lengths, queue discharge headways, start-up lost times, volumes, speeds, control delays and stop times, lane-specific queue lengths, and signal states. The two scenarios also made use of aggregate vehicle and bus arrivals detected by detectors (the “upstream detectors”), placed 100 to 200 ft from the upstream end of the link. Note that these are the vehicle and bus arrival detectors described in Section 6.2.3. In addition, Scenario 1 also made use of lane specific occupancy and speed detected by “downstream detectors” placed on each lane, 120 ft upstream of the stop line. Thus, Scenario 1 had 30 input units and 24 output units, while Scenario 2 had 18 input units and 12 output units.

Table 6.2 **Traffic Parameters Used in Each Scenario for Arterial Flow Prediction and Traffic State Prediction Modules**

Traffic Data	Scenario 1	Scenario 2
Input Data		
Link length	x	x
Queue discharge headway	x	x
Start-up lost time	x	x
Total vehicle arrivals detected by upstream detectors at time t	x	x
Total bus arrivals detected by upstream detectors at time t	x	x
Occupancy detected by downstream detector on lane 1 (2, 3, ..6) at time t	x	
Speed detected by downstream detector on lane 1(2, 3, ..6) at time t	x	
Link volume at time t	x	x
Link speed at time t	x	x
Signal state for through movement from the link at time t	x	x
Signal state for right-turn movement from the link at time t	x	x
Signal state for left-turn movement from the link at time t	x	x
Number of vehicles in queue on lane 1 (2, 3,..6) at time t	x	x
Control delay at time t	x	x
Stop times at time t	x	x
Output Data		
Total vehicle arrivals detected by upstream detectors at time t+1	x	x
Total bus arrivals detected by upstream detectors at time t+1	x	x
Occupancy detected by downstream detector on lane 1 (2, 3, ..6) at time t+1	x	
Speed detected by downstream detector on lane 1(2, 3, ..6) at time t+1	x	
Link volume at time t+1	x	x
Link speed at time t+1	x	x
Number of vehicles in queue on lane 1 (2, 3,..6) at time t+1	x	x
Control delay at time t+1	x	x
Stop times at time t+1	x	x

The next step was to determine the optimum number of hidden units. There are a number of “rules of thumb” cited in the literature in choosing the optimum number of hidden units. However, deciding the network architecture, including the number of hidden units is a trial and error process, since the optimum architecture depends on the number of training patterns, the noise in the data, the extent of generalization that is needed, etc. (Haykin, 1994). Hence, various architectures were considered for the two scenarios by varying the number of hidden units systematically from 2 to 20. The hidden layers were varied from 1 to 2.

Data were normalized between 0.25 and 0.85, and training was done using the Qnet 2000 software tool (1999), which makes use of the backpropagation training algorithm. The sigmoid function was chosen as the activation function. For Scenario 1, the best architecture made use of 1 hidden layer and 9 hidden units, while for Scenario 2, the best architecture made use of 1 hidden layer with 5 hidden units. Table 6.3 presents the learning errors (training and validation errors) of the two model structures. The results show that neither scenario outperformed the other. The differences between the two were not statistically significant. However, given that Scenario 1 was more data intensive and required the presence of additional detectors at a location of 120 ft, the input structure of Scenario 2 was chosen for this study. The Scenario 2 model was then applied to the test data set to estimate the generalization errors, summarized in Table 6.4. The results show that the chosen model was able to predict the output patterns within an acceptable range of accuracy.

Table 6.3 Comparison of Learning Errors for the Two Model Structures

MOE	Scenario 1		Scenario 2	
	Training Error	Validation Error	Training Error	Validation Error
Number of vehicles queued to make a through or a right-turn movement	2.66	2.03	2.85	2.1
Number of vehicles queued to make a left-turn movement	1.44	1.27	1.48	1.21
Control delay	2.36	1.77	2.53	1.83
Stop times	2.64	2.09	2.92	2.21
Link volume	16.54	16.63	17.13	15.41
Link speed	10.76	11.53	9.58	10.2

Table 6.4 Generalization Errors for the ANN I Model

MOE	Generalization Error	
	RMS	Mean Percentage Error
Number of vehicles queued to make a through or a right-turn movement	2.7	11.88%
Number of vehicles queued to make a left-turn movement	1.42	9.50%
Control delay	2.65	13.05%
Stop times	2.63	9.67%
Link volume	18.36	18.51%
Link speed	9.73	11.53%

Developing the ANN Model for Bus Prediction Module

The bus prediction module, described in Section 5.2.1, predicts the time a bus would join the queue or reach the stop bar to compete for priority treatment. To develop the required model, ANN II, three model structures were examined. These are summarized in Table 6.5. For links that did not have a bus station, or alternately, when a bus was not scheduled to make a stop on a link, the time when the bus was detected by the vehicle and bus arrival detector (upstream detectors) was taken as time t.

Table 6.5 Traffic Parameters Used in Each Scenario for Bus Prediction Module

Traffic Data	Scenario 1	Scenario 2	Scenario 3
Input Data			
Time t when bus clears the bus stop	x	x	x
Location of bus from the stop bar at time t	x	x	x
Speed of bus at time t	x	x	x
Next turn movement of the bus	x	x	
Queue discharge headway	x		x
Start-up lost time	x		x
Link volume at time t	x	x	x
Link speed at time t	x		x
Occupancy detected by downstream detector on lane 1 (2, 3, ..6) at time t	x	x	
Speed detected by downstream detector on lane 1(2, 3, ..6) at time t	x	x	
Signal state for through movement from the link at time t	x	x	x
Signal state for right-turn movement from the link at time t	x	x	x
Signal state for left-turn movement from the link at time t	x	x	x
Number of vehicles in queue on lane 1 (2, 3,..6) at time t	x	x	x
Output Data			
Occupancy detected by downstream detector on lane 1 (2, 3, ..6) at time t+1	x	x	
Speed detected by downstream detector on lane 1(2, 3, ..6) at time t+1	x	x	
Speed of bus at time t+1	x	x	x
Time when bus will join queue or reach the stop bar	x	x	x

Scenarios 1 and 2 had 29 and 26 input units, respectively. Both scenarios had 14 output units. Scenario 3 had 16 input units and 2 output units. To determine the optimum hidden units, the performances of the three models were examined by varying the number of hidden units from 2 to 25, and the hidden layers from 1 to 2. Training was performed using Qnet. For Scenarios 1 and 2, the best architecture made use of 7 hidden units, while for Scenario 3 the optimum hidden unit was determined to be 4. Table 6.6 compares the training and validation errors of the best architectures for the three scenarios. The training and validation errors for Scenario 1 were higher than those for Scenarios 2 and 3. As Scenario 2 made use of additional detectors, Scenario 3 was chosen for predicting the bus speed at the next control time interval and the time when a bus would join the queue. The test data set was applied to the Scenario 3 model to estimate the generalization errors (Table 6.7). As can be seen from the Table 6.7, the generalization errors were very low.

The weights of the two ANN models developed in this section, is used to predict arterial and side-street flows and bus data, which is used to determine signal control decisions.

Table 6.6 Comparison of Learning Errors of the Bus Data Prediction Models

MOE	Scenario 1		Scenario 2		Scenario 3	
	Training Error	Validation Error	Training Error	Validation Error	Training Error	Validation Error
Time when bus will join queue	1.12	1.79	0.39	0.39	0.34	0.37
Speed of bus	3.23	3.17	2.29	2.07	2.32	2.12

Table 6.7 Generalization Errors for ANN II Model

MOE	Generalization Error	
	RMS	Mean Percentage Error
Time when bus will join queue	0.41	0.71%
Speed of bus	2.56	5.20%

6.4 DEVELOPMENT OF THE MODELS USING CORSIM RUN-TIME EXTENSION

CORSIM's run-time extension feature was used to access the data structures in CORSIM every second, and to modify the signal states as determined by the models developed in Chapters 4 and 5 while executing the simulation. The arterial signal control models were written in Visual C++. Note that alternative (i) described in Section 6.2.4 did not require the use of CORSIM's run-time extension feature since CORSIM is capable of modeling actuated signal control.

6.4.1 Extraction of Required Data

Using the run-time extension, vehicle and bus arrivals, vehicle queue lengths, arterial volumes, and signal states were extracted every second by accessing the relevant CORSIM data structures. Vehicle arrivals for each link were determined by keeping track of the vehicles that activated the vehicle and bus arrival detector placed at the upstream end of the link. If the vehicle activating the detector was determined to be a bus, then the bus arrivals for the link were increased by one. Vehicle and bus arrivals thus detected were accumulated for a period of five seconds (the duration of one control time interval).

To compute lane-specific vehicle queue lengths for each link, a check was performed to see whether the first vehicle on the lane was within the intersection control length. In this study intersection control length was determined as the minimum distance required by a vehicle traveling at its mean desired speed to come to a stop. If it was determined that the first vehicle was under intersection control, then the vehicle queue length was incremented by one. A follower vehicle was said to be "in queue" if it was not a bus stopped at a bus stop and was either stopped or traveling at a speed less than 10 fps and an acceleration of less than 2 fps^2. All vehicles on the lane were scanned systematically to see if they were in queue. If it was determined that a vehicle following another vehicle that was already determined to be in queue, was also in queue, the vehicle queue was incremented by one. Whenever, it was identified that a vehicle was not in queue, accumulation of vehicle queue for the lane was stopped. Note that if the first vehicle on the lane was not determined to be under intersection control, the vehicle queue for the lane was set to be zero. Vehicle queues thus determined were averaged over a period of five seconds.

Signal states for the previous second were extracted from CORSIM's data structure. Link volumes, speeds, control delays, and stop times for the previous second were directly extracted from the CORSIM output.

Data thus extracted using the run-time extension were used not only in developing the neural network models but also in developing the arterial signal control models for determining the optimal signal control decisions.

6.4.2 Modeling Progression Control Level

For Model II, Model III, and Model IV, the bandwidth maximization problem described in Section 4.3 was formulated as a mixed-integer programming problem in MPL 4.2 (2004), which uses CPLEX to solve the optimization problem. MPL 4.2 has the capability of allowing external applications to call it directly to solve an optimization problem. Using this feature, it was possible to solve the bandwidth maximization problem by calling MPL and running it in run-time mode while executing the arterial signal control models.

Using the run-time extension, data required by the bandwidth maximization problem, such as vehicle and bus arrivals, vehicle queue lengths, and arterial volumes were extracted every second. Real-time data were used only for the first control time interval in the time horizon, specified to be equal to the system cycle length. For Model III and Model IV, which used robust optimization, the arterial volumes for the first control time interval were varied from -30% to 30% of the detected volumes to capture possible fluctuations in traffic. For Model II, the arterial volumes were not modified. Since the bandwidth maximization problem was solved over the entire time horizon, data for subsequent control time intervals in the time horizon were predicted using ANN I developed for arterial flow prediction in Section 6.3.

For MULTIBAND+ and Model I, the MULTIBAND method was formulated in MPL 4.2. To emulate the MULTIBAND method, real-time arterial volumes detected for the first control time interval were kept a constant for solving the bandwidth maximization problem. Volumes were not predicted or estimated for subsequent time intervals. Queue clearance time was pre-specified as 40 seconds for arterial links and 20 seconds for side street links. Minimum green was specified to be 20 seconds for all links.

The bandwidth maximization problem returned the optimal system cycle length. The progression control was called only once every cycle length. At the beginning of the control time interval, if a bus was detected that warranted priority consideration, then the bus priority optimization method was called. Otherwise, the intersection control was called.

6.4.3 Modeling Intersection Control Level

The intersection control was called once every five seconds (duration of the control time interval). For Model I, Model II, Model III and Model IV, the intersection optimization problem was solved for the entire time horizon using the approach described in Section 4.4.3. As observed earlier, since real-time data were used only for the first control time interval, ANN I was used to predict vehicle queue lengths, control delays, and stop times for subsequent time intervals in the time horizon to compute the transformation functions.

For MULTIBAND+, the intersection optimization problem was solved for only one control time interval using real-time data.

Signal states that minimized the performance index were stored for each of the four intersections. Note that this study assigned equal weights to vehicle queues, stop times and control delays.

6.4.4 Modeling Bus Priority Control Level

Bus priority was included only in Model IV. As was done at the intersection control level, real-time data were used for control delays and queue lengths, for the first control time interval. For subsequent time intervals, ANN I was used to predict the queue lengths and control delays. Bus data such as bus speed and time when a bus would join the queue were determined using ANN II, developed in Section 6.3. If it was determined that a detected bus would join the queue, the schedule delay for the bus was also included in the performance index. The performance index was computed as a sum of the control delays, schedule delays, and automobile and bus passenger delays over all control time intervals in the time horizon. Automobile passenger delays were computed by multiplying vehicle queue lengths by the average occupancy of automobiles, assumed to be 2 in this study. Bus passenger delays were computed by adding the occupancy of detected buses that were in queue or were predicted to join the queue within the control time interval. Signal states that minimized the performance index over the time horizon were determined.

The signal states for the four intersections were specified every second based on the signal control decisions, determined for the first control time interval (five seconds), by overwriting the relevant signal control data structures in CORSIM.

For this case study, the total execution time was 8 minutes for Model I, 10 minutes for Model II, and 16 minutes each for Model III and Model IV, on a laptop computer with Intel Pentium 1.8 GHz Processor with 512 MB RAM. Thus, the execution time for accessing the data structures every second, which was used by all three levels, was 25% of the total execution time. The individual execution times for the intersection control level, the progression control level without robust optimization, and the progression control level with robust optimization, were 25%, 12.5% and 37.5%, respectively, of the total execution time. The execution times for Model III and Model IV were identical, implying that no additional computational time was needed in excess of what was required by the intersection control level. This is because the bus priority control level was called only when a bus was detected by the surveillance system; otherwise, the signal timing plans were determined by the intersection control level.

6.5 PERFORMANCE EVALUATION RESULTS

To evaluate the incremental benefits of the models developed at the three control levels, the six signal control alternatives described in Section 6.2.4 were compared by applying them to the traffic scenario presented in Table 6.1. The measures of effectiveness described in Section 6.2.5 were accumulated every 60 seconds over a period of one hour for each of the models. Statistics for the first 600 seconds were disregarded. Figures 6.2 to 6.5 illustrate the average control delays, stop times, queues, and bus delays for every minute for an arterial link. Figures 6.6 to 6.9 present the corresponding measures for a side street link. Tables 6.8 and 6.9 summarize the average performance measures for arterial and side-street links over the simulation period.

6.5.1 Performance Evaluation of the Proposed Intersection Control Model

To analyze the effectiveness of the intersection control model described in Chapter 4, Model I was compared with the Baseline Model and MULTIBAND+ (Tables 6.8 and 6.9). Figures 6.10 and 6.11 show that although Model I performed better than the other two alternatives in reducing control delays, stop times and vehicle queues, and increasing link speeds, the benefits were nominal compared to MULTIBAND+, especially for arterial links. Although Model I optimized vehicle queues, control delays and stop times over the entire time horizon using predicted traffic flows, the signal decision was constrained by bandwidths provided by the MULTIBAND method, which used static flows and pre-specified queue clearance times. Hence, Model I was limited by its progression control model, which is deficient in real-time applications.

6.5.2 Performance Evaluation of the Proposed Progression Control Model without Robust Optimization

To evaluate the incremental benefits of the progression control model described in Chapter 4 without robust optimization, Model II was compared with the Baseline Model, MULTIBAND+ and Model I. Model II demonstrated superior performance over the Baseline Model, MULTIBAND+ and Model I (Tables 6.8 and 6.9). Compared to Model I, Model II reduced control delays by 12%, stop times by 14% and queues by 11% for arterial links (Figure 6.12). Speeds increased by 9%. Benefits for side street links were lower and ranged from 5% to 7% (Figure 6.13). With respect to the performance of MULTIBAND+, the benefits from Model II were significantly higher. Model II resulted in a 12% decrease in control delays, 15% decrease in stop times and 12% decrease in vehicle queues for arterial links. Link speeds increased by 9%. For side street links, Model II resulted in benefits ranging from 8% to 11%. Compared to the Baseline Model, the benefits were higher ranging from 8% to 24%.

The total throughput along the arterial was found to be 20114 vehicles for the Baseline Model, 20346 vehicles for MULTIBAND+, and 20397 vehicles for the Model I. With Model II, it increased to 20988 vehicles. Integrating the intersection control model with the proposed progression control model that eliminated the limitations of the MULTIBAND method proved to be highly beneficial.

6.5.3 Performance Evaluation of the Proposed Progression Control Model with Robust Optimization

To analyze the effectiveness of the robust discrete optimization technique, the benefits of the arterial signal control model that employed robust optimization (Model III) was compared with the model that did not use such a technique (Model II), MULTIBAND+, Model I and the signal control in the field (Baseline Model). Figures 6.14 to 6.15 show that overall, Model III showed the best performance with respect to control delay, stop times, and queue lengths for both arterial as well as side street links.

Figures 6.2 to 6.7 show that Model II demonstrated higher benefits for arterial links during the initial 20 minutes when traffic volumes were low. As traffic volumes increased, Model III showed better performance. This was as expected, since benefits from robust optimization increase with increase in traffic volume and arrival flow variability. There was a nominal increase in average link speeds by 2%.

With respect to MULTIBAND+, Model III demonstrated superior performance. For arterial links, Model III reduced control delays by 17%, stop times by 20%, and vehicle queues by 18%. Compared to the Baseline Model, the corresponding benefits were a 21% reduction in control delays, 20% reduction in stop times and 29% reduction in vehicle queues. Comparable benefits were also observed for side street links.

Moreover, with Model III arterial throughput was 21199 vehicles. Thus, including robust optimization technique in the arterial signal control model proved to be highly beneficial in providing arterial progression as well as reducing system-wide delays and stop times.

6.5.4 Performance Evaluation of the Bus Priority Model

To evaluate the benefits of bus priority, bus performance measures, including bus delay time and bus speed were compared for all six signal control alternatives. Figures 6.5 and 6.9 illustrate that Model IV produced significant reductions in bus delays. Note that the statistic depicted in the charts are the average delays experienced by a bus averaged over all arterial or side street links.

From Figures 6.16 and 6.17 it can be deduced that compared to Model III, Model IV was able to reduce bus delays by 19% on arterial links and by 13% on side street links. Compared to the Baseline Model, Model IV produced reductions of 31% and 36% in average bus delays on arterial and side street links. Reductions with respect to the MULTIBAND+ model were 32% for arterial links and 40% for side street links.

Model IV also resulted in increased bus speeds. With respect to Model III, the bus speeds increased by 5% for arterial links and 2% for side street links. With respect to the Baseline Model and the MULTIBAND+ model, the benefits were higher and ranged from 19% to 30%. However, Model IV did not outweigh the benefits of Model III in reducing control delays, stop times and vehicle queues. This was as expected since Model IV also included priority treatment to buses in its signal optimization function. Model II and Model III were able to realize significant benefits for buses even though their objective was not to give priority to buses. Moreover, despite providing priority to buses, Model IV performed better than the Baseline Model and the MULTIBAND+ with respect to reduction in automobile-disutility. Compared to the Baseline Model, Model IV reduced control delays by 18%, stop times by 16% and queues by 24%, and increased speeds by 10% for arterial links. With respect to MULTIBAND+, Model IV resulted in reductions of 13%, 16% and 12% in control delays, stop times, and vehicle queues, and increase of 10% in speeds for arterial links. Side street links also experienced higher benefits with Model IV. With Model IV, the total arterial throughput was 20926 vehicles. Thus, this research was able to produce a model that gave priority to buses without causing significant delays to automobiles.

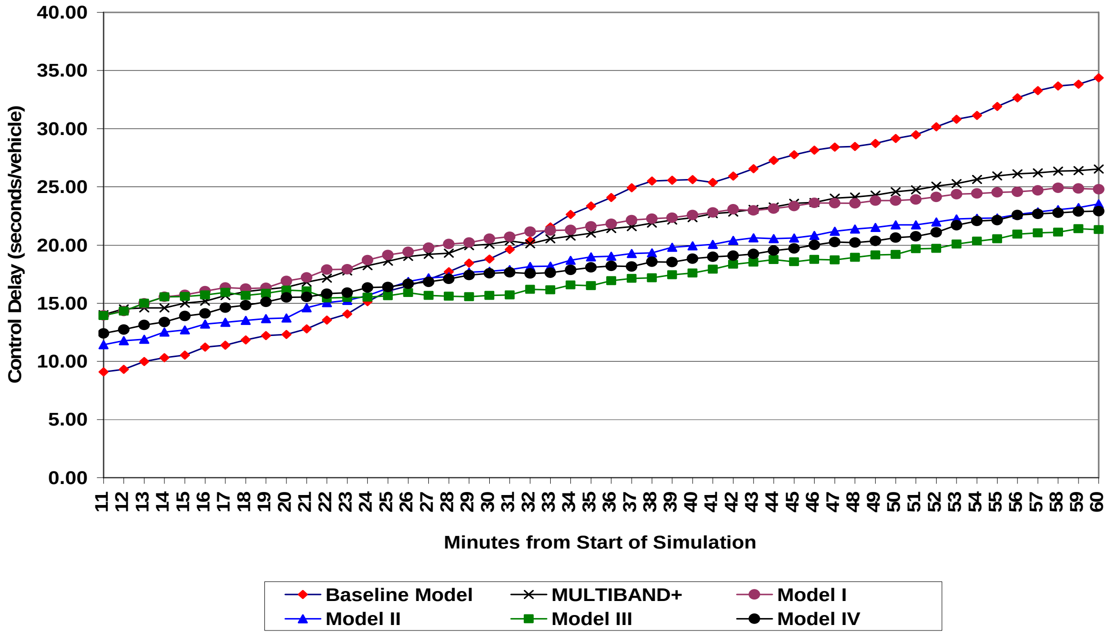

Figure 6.2 Average Control Delay on an Arterial Link

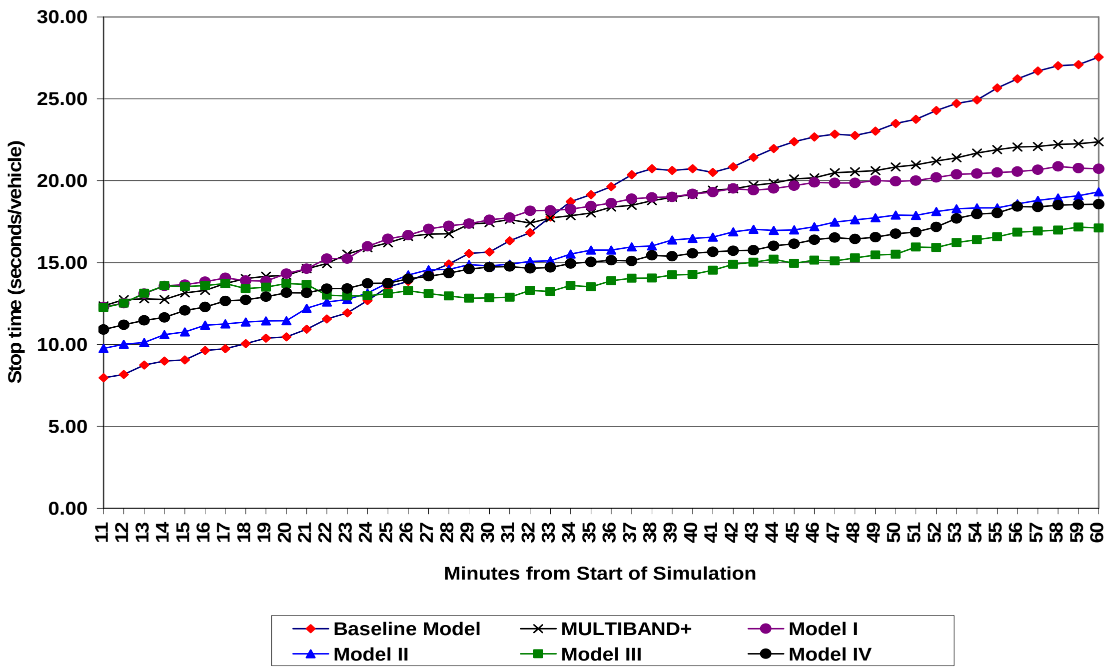

Figure 6.3 Average Stop Times on an Arterial Link

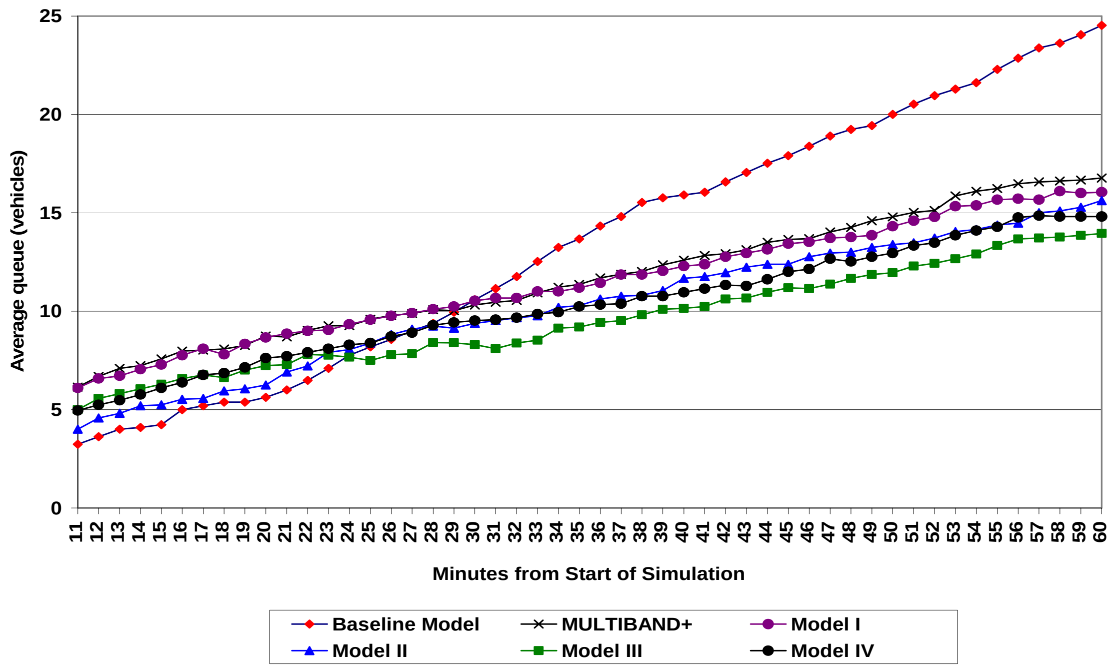

Figure 6.4 Average Queue on an Arterial Link

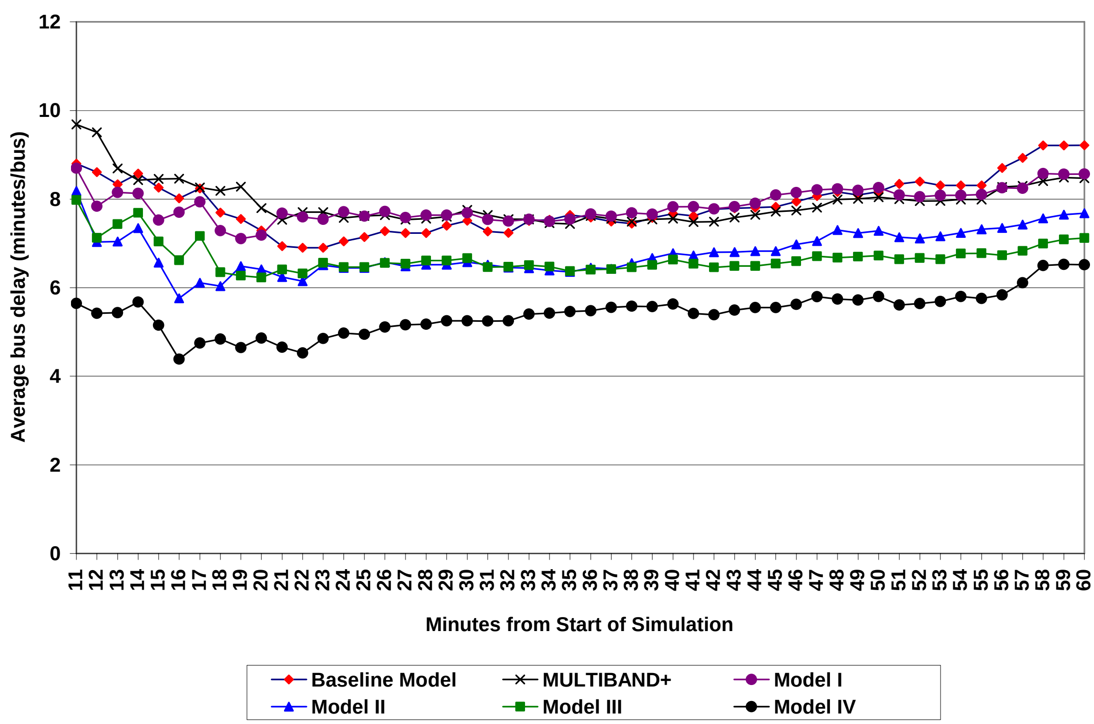

Figure 6.5 Average Bus Delay on an Arterial Link

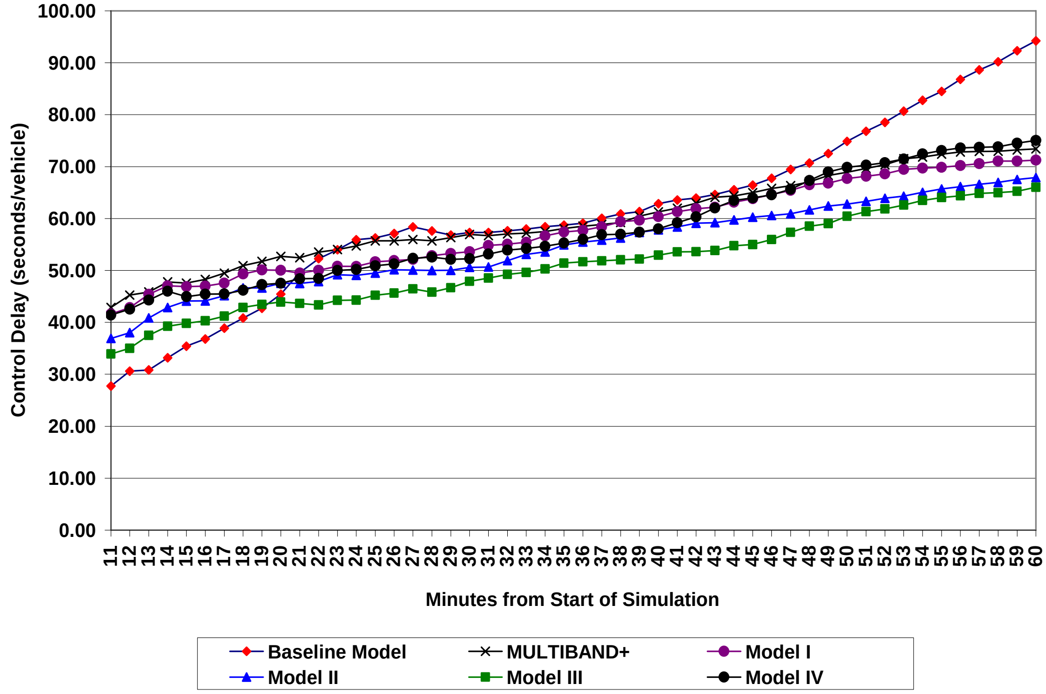

Figure 6.6 Average Control Delay on a Side Street Link

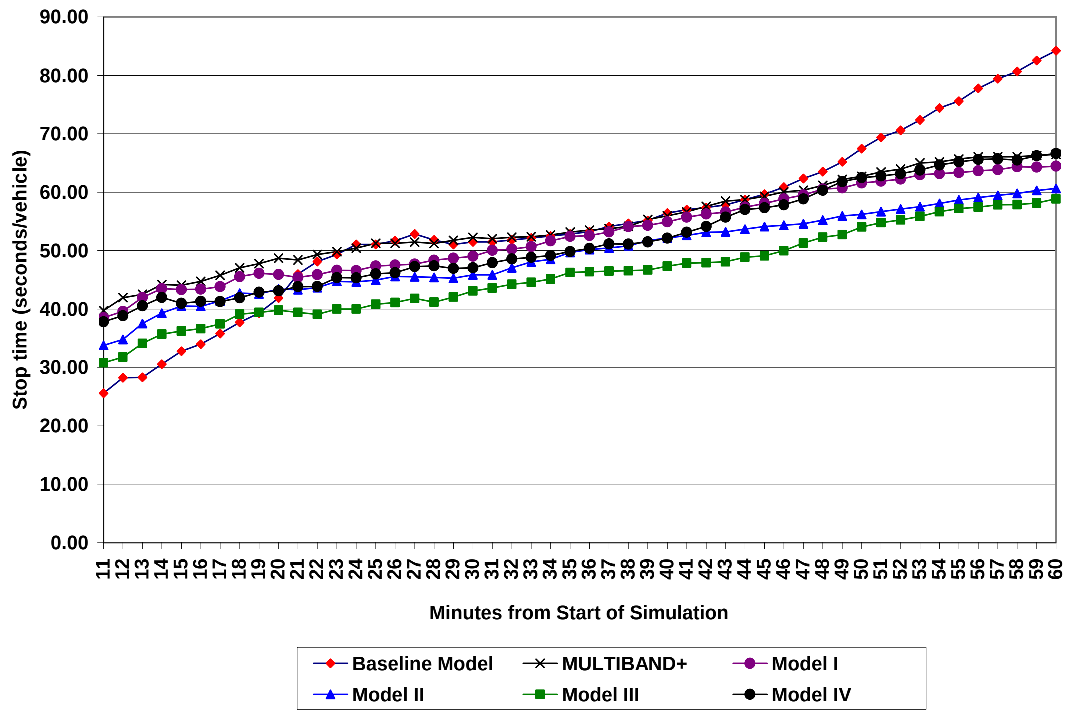

Figure 6.7 Average Stop Times on a Side Street Link

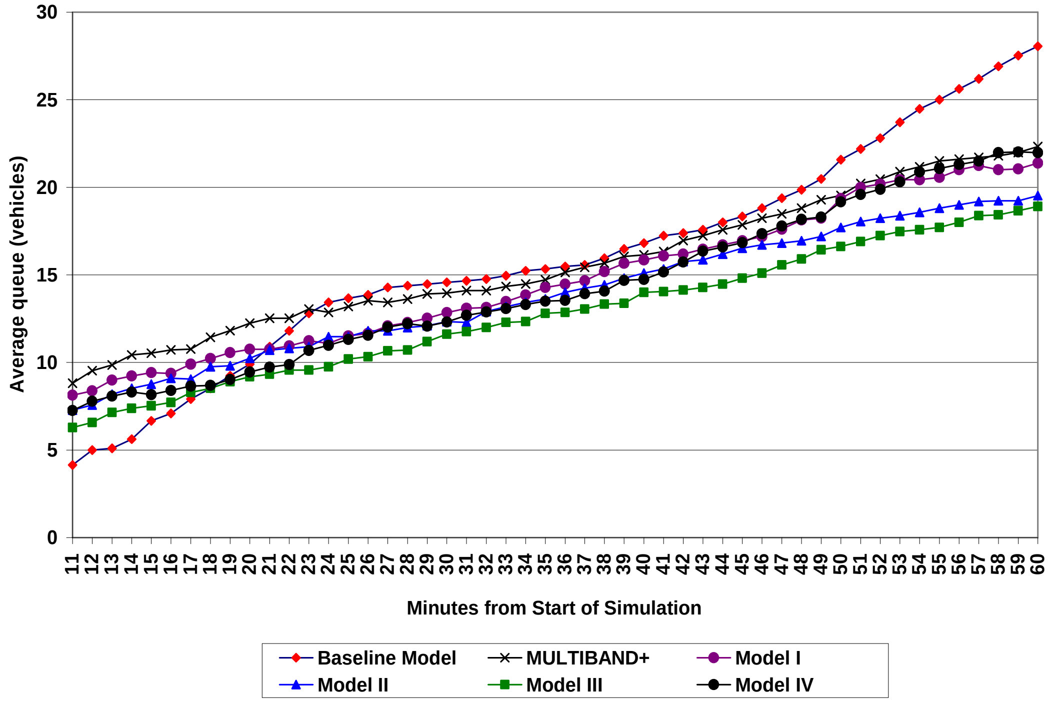

Figure 6.8 Average Queue on a Side Street Link

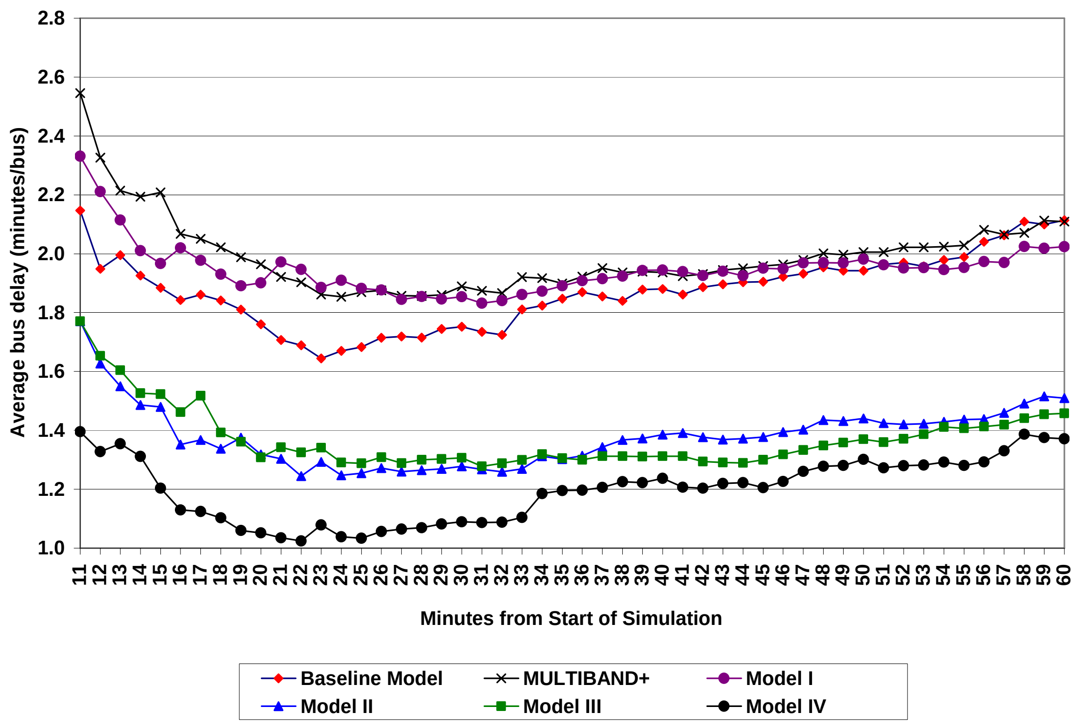

Figure 6.9 Average Bus Delay on a Side Street Link

Table 6.8 Average Performance Measures for an Arterial Link over the Simulation Period

MOE	Baseline Model		MULTIBAND+		Model I		Model II		Model III		Model IV	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Control Delay (seconds/vehicle)	21.97	7.95	20.90	3.84	20.87	3.30	18.31	3.57	17.41	2.07	18.12	2.90
Stop Time (seconds/vehicle)	17.98	6.12	17.94	3.07	17.73	2.62	15.21	2.83	14.36	1.40	15.06	2.09
Queue Length (vehicles)	13.47	6.66	11.71	3.11	11.48	2.91	10.26	3.31	9.56	2.54	10.29	2.87
Link Speed (mph)	22.44	1.66	22.49	0.83	22.50	0.72	24.54	0.82	24.99	0.81	24.64	0.92
Bus Delay Time (minutes/bus)	7.88	0.62	7.93	0.49	7.86	0.36	6.80	0.49	6.68	0.35	5.43	0.47
Bus Speed (mph)	15.74	1.15	15.40	1.33	15.95	1.20	17.80	1.29	17.84	1.29	18.75	1.36

Table 6.9 Average Performance Measures for a Side Street Link over the Simulation Period

MOE	Baseline Model		MULTIBAND+		Model I		Model II		Model III		Model IV	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Control Delay (seconds/vehicle)	60.94	16.78	59.89	8.56	58.06	8.67	54.73	8.28	51.23	8.74	57.81	10.18
Stop Time (seconds/vehicle)	54.99	14.72	54.80	7.46	52.97	7.60	49.40	7.05	45.97	7.57	51.97	8.78
Queue Length (vehicles)	15.99	6.29	15.66	3.86	14.72	4.10	13.90	3.64	12.78	3.68	14.30	4.56
Link Speed (mph)	15.79	2.13	15.86	1.10	16.28	1.15	17.06	1.14	17.64	0.84	16.81	1.10
Bus Delay Time (minutes/bus)	1.87	0.13	1.99	0.13	1.95	0.09	1.38	0.10	1.37	0.10	1.20	0.11
Bus Speed (mph)	10.04	3.47	9.65	3.79	9.77	3.66	12.23	3.55	12.29	3.61	12.53	3.64

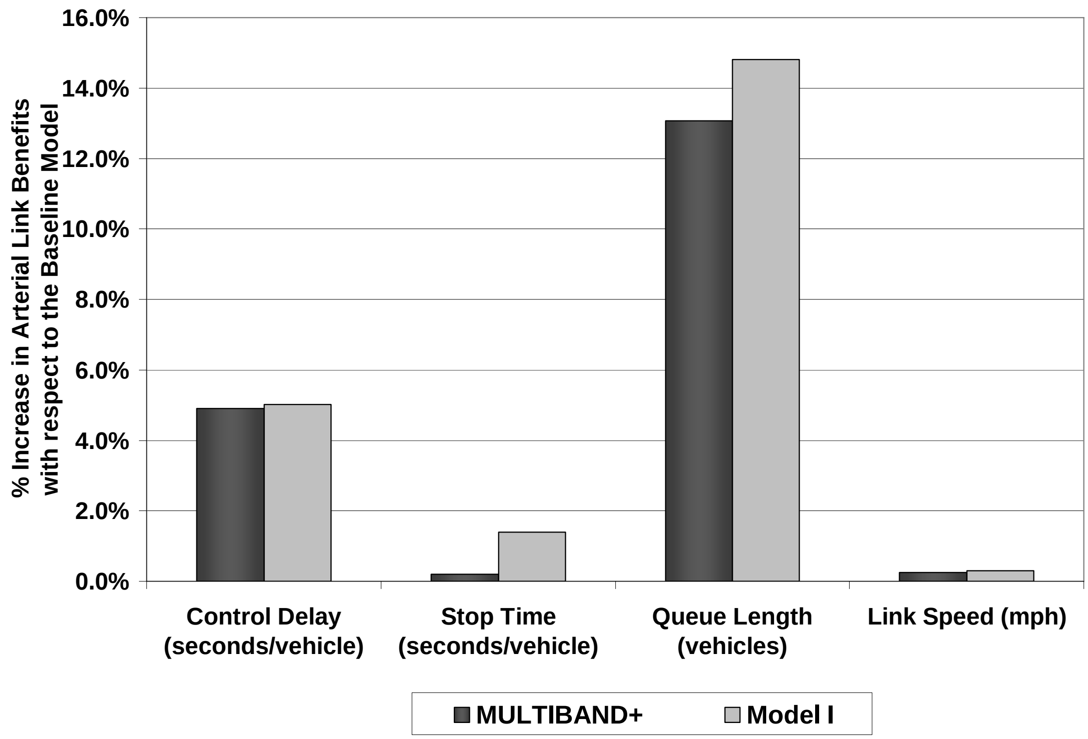

Figure 6.10 Arterial Link Benefits with Intersection Control Level Model

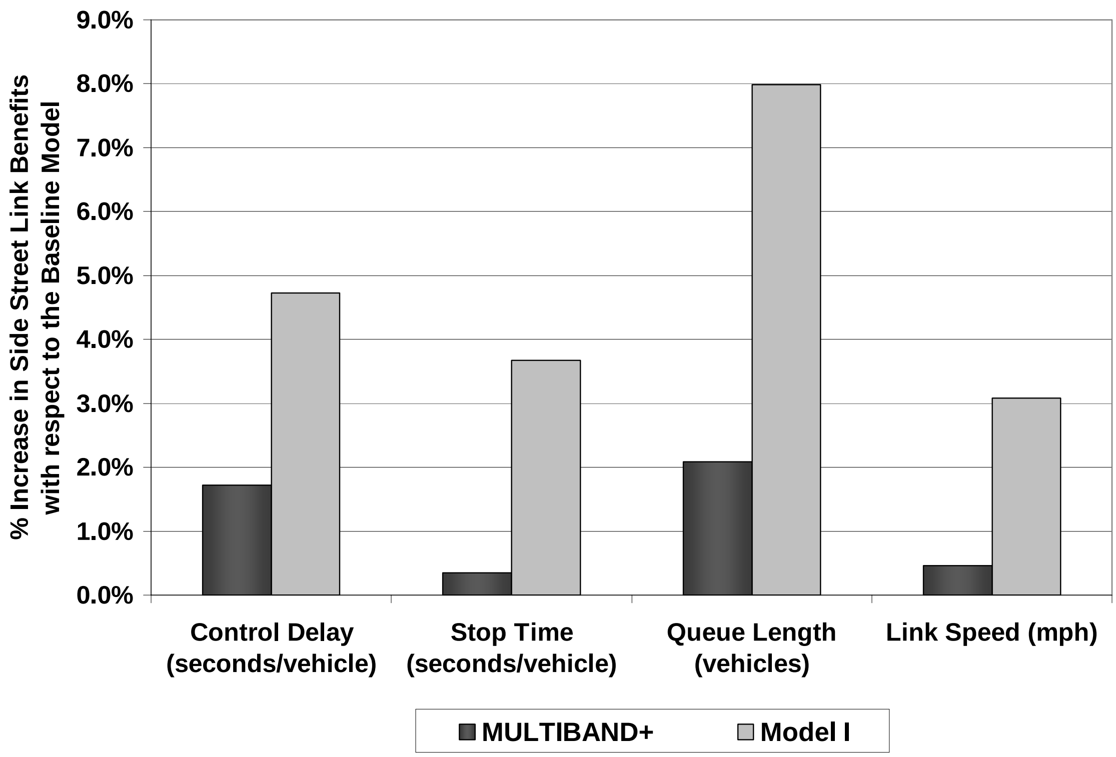

Figure 6.11 Side Street Link Benefits with Intersection Control Level Model

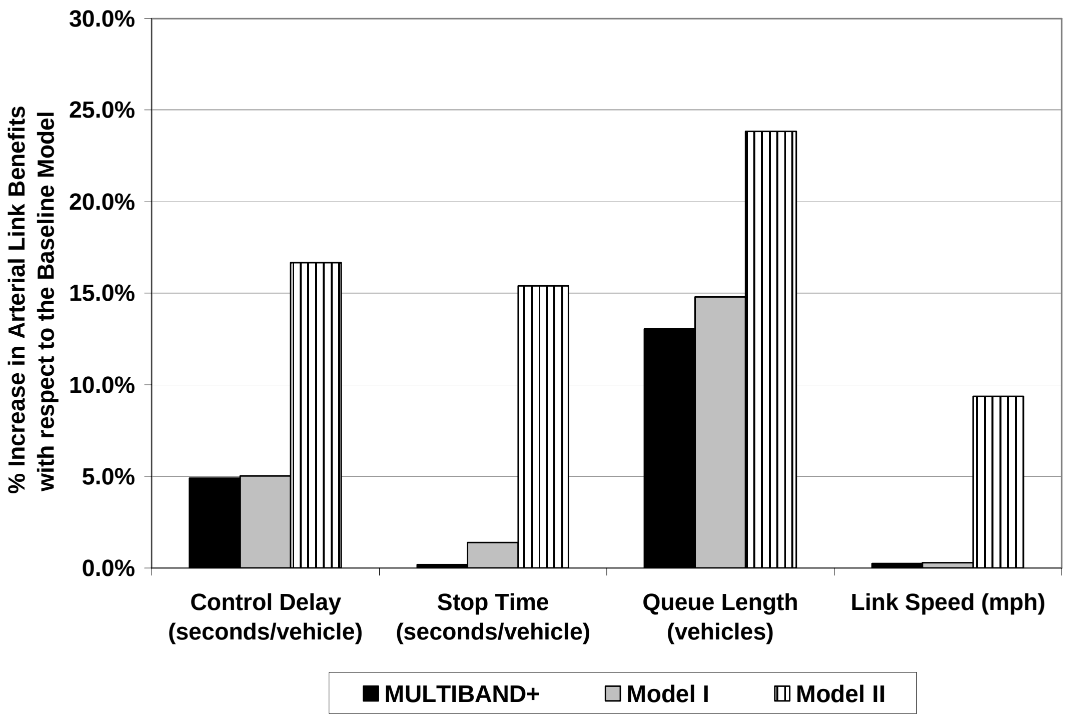

Figure 6.12 Arterial Link Benefits with Progression Control Level Model without Robust Optimization

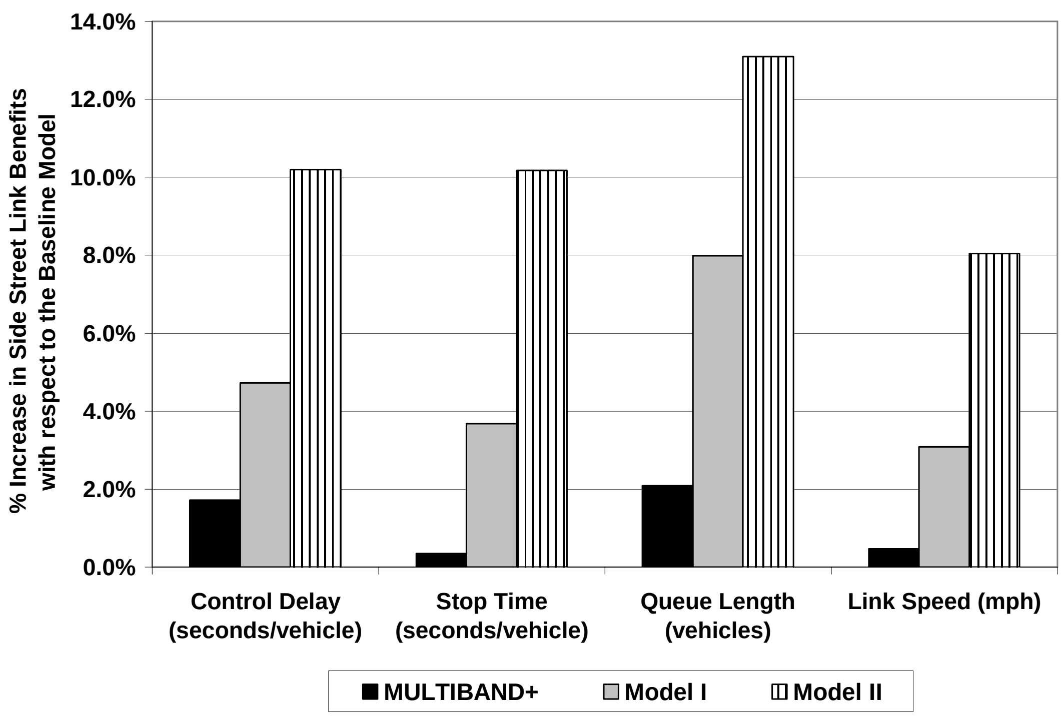

Figure 6.13 Side Street Link Benefits with Progression Control Level Model without Robust Optimization

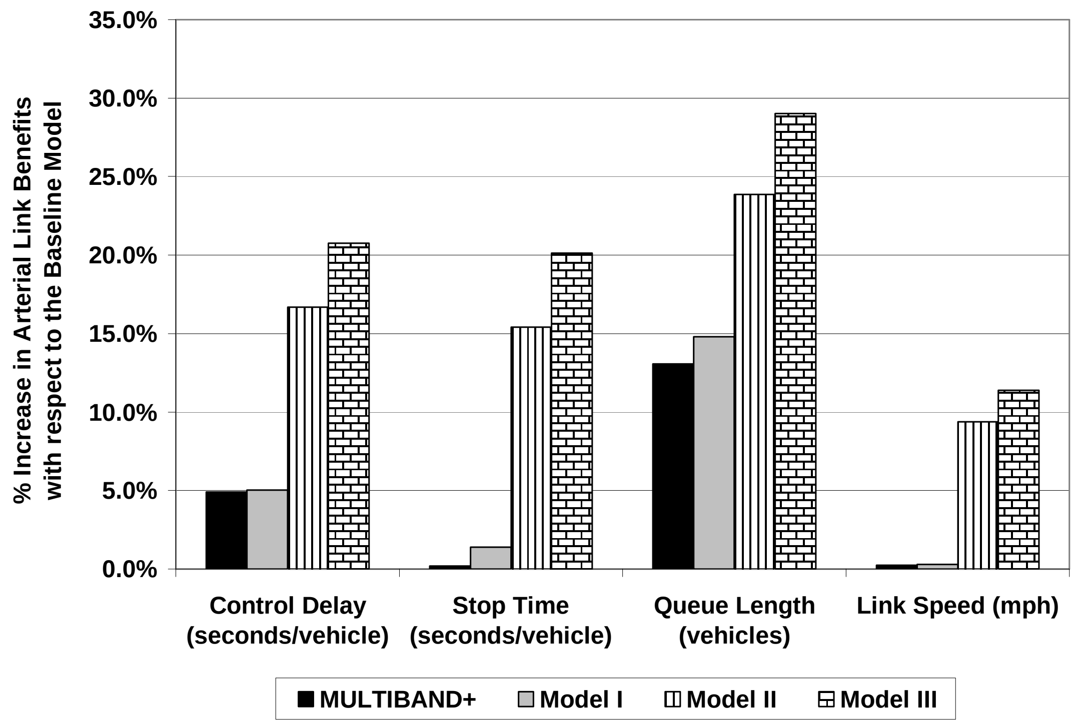

Figure 6.14 Arterial Link Benefits with Progression Control Level Model with Robust Optimization

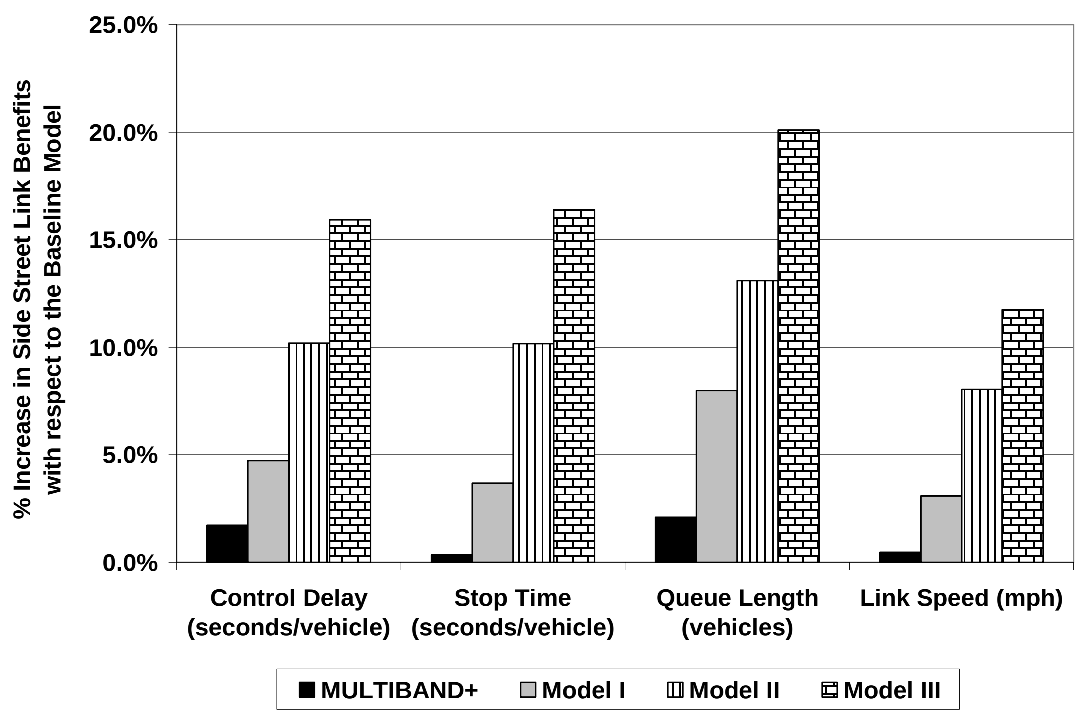

Figure 6.15 Side Street Link Benefits with Progression Control Level Model with Robust Optimization

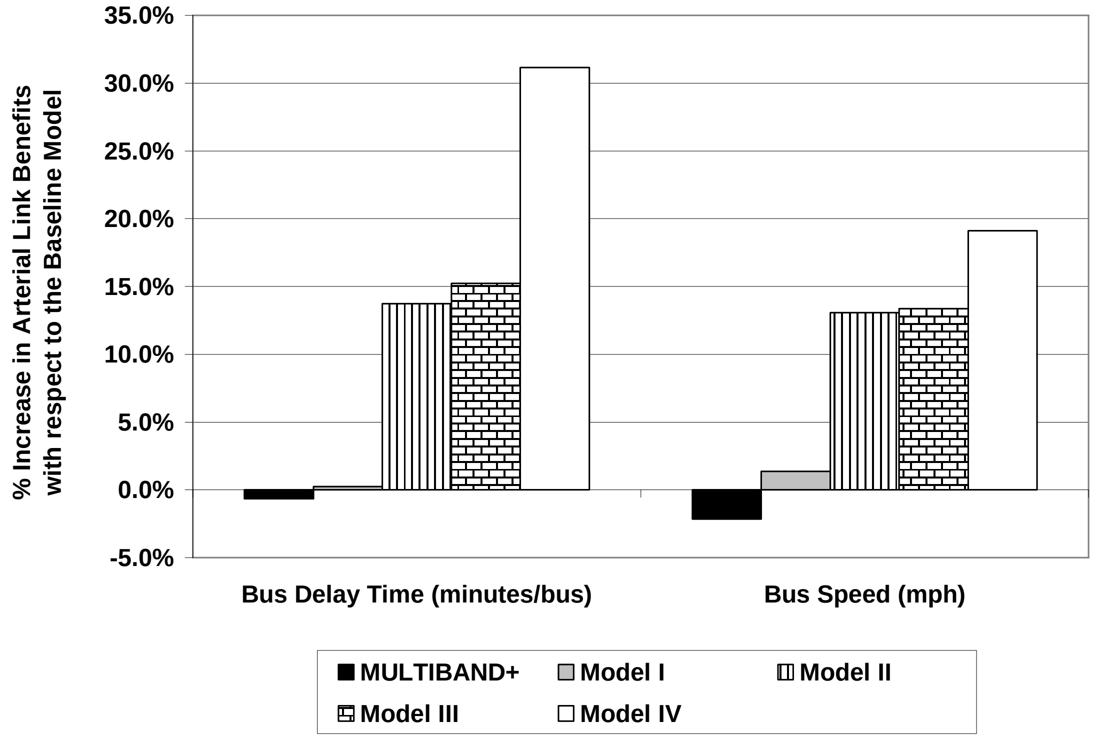

Figure 6.16 Arterial Link Benefits with Bus Priority Control Level Model

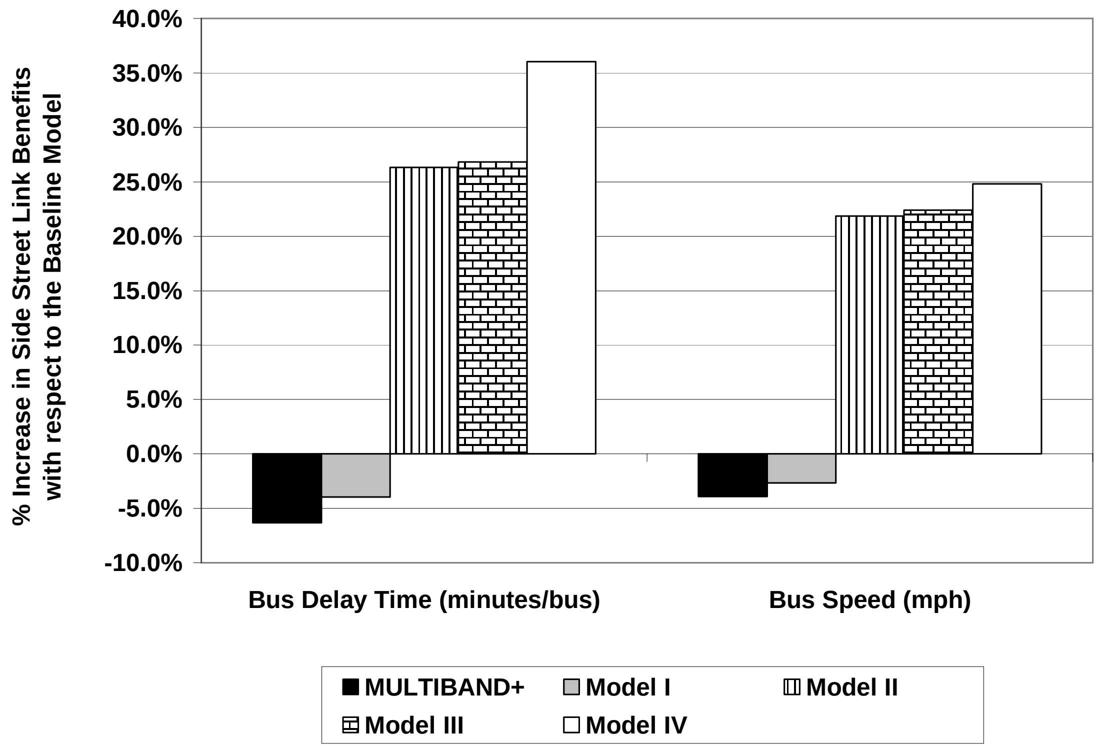

Figure 6.17 Side Street Link Benefits with Bus Priority Control Level Model

6.5.5 Sensitivity Analysis of Traffic Volume Variation on Benefits of Robust Optimization

Section 6.5.3 demonstrated that when traffic volumes increased, the performance of Model III, which used robust optimization technique improved. To further evaluate the effectiveness of robust optimization, this section compares the performance of the Baseline Model, MULTIBAND+, Model I, Model II and Model III by varying traffic volumes by δ% of the field data to reproduce the best and the worst realizable scenarios. Six scenarios were modeled by varying δ% as 5%, 10%, 15%, 20%, 25%, and 30%. Traffic volumes were specified in the CORSIM input file at 15-minute intervals. The first 15-minute interval used the field volumes. For subsequent 15-minute intervals, traffic volumes were alternately increased and decreased by δ%.

Figures 6.18 to 6.20 illustrate the performance measures for the different scenarios for arterial links. Performance of Model III outweighed those of the other signal control alternatives for all levels of traffic volumes. As can be inferred from the charts, when traffic volume variation increased, Model III demonstrated significant benefits. When the traffic volumes were varied by 30%, Model III reduced control delays, stop times and vehicle queues by more than 35% compared to the Baseline Model. Compared to the alternative that did not use robust optimization technique (Model II), the reductions ranged from 9% to 13%. With respect to the alternatives that used the MULTIBAND method for signal progression (MULTIBAND+ and Model I), the corresponding benefits ranged from 20 to 31%. Thus, incorporating robust optimization to solve the signal progression problem produced significant benefits when there were fluctuations in traffic flow patterns.

6.5.6 Sensitivity Analysis of Bus Data

In this study, since no AVL/CAD information was available, bus schedule delays and bus occupancy were assumed to follow normal distributions. In order to validate the simulation results, it was essential to perform a sensitivity analysis. The means and the standard deviations of the normal distributions were varied and the performance of the six models tested.

The benefits of Model IV, which gave special treatment to buses, with respect to the other signal control alternatives are summarized in Tables 6.10 to 6.13. From the tables, it can be seen that performance of Model IV was similar to what was observed earlier (Tables 6.8 and 6.9). Model IV resulted in significant reductions in bus delay and increase in bus speed compared to the other alternatives for all parameters.

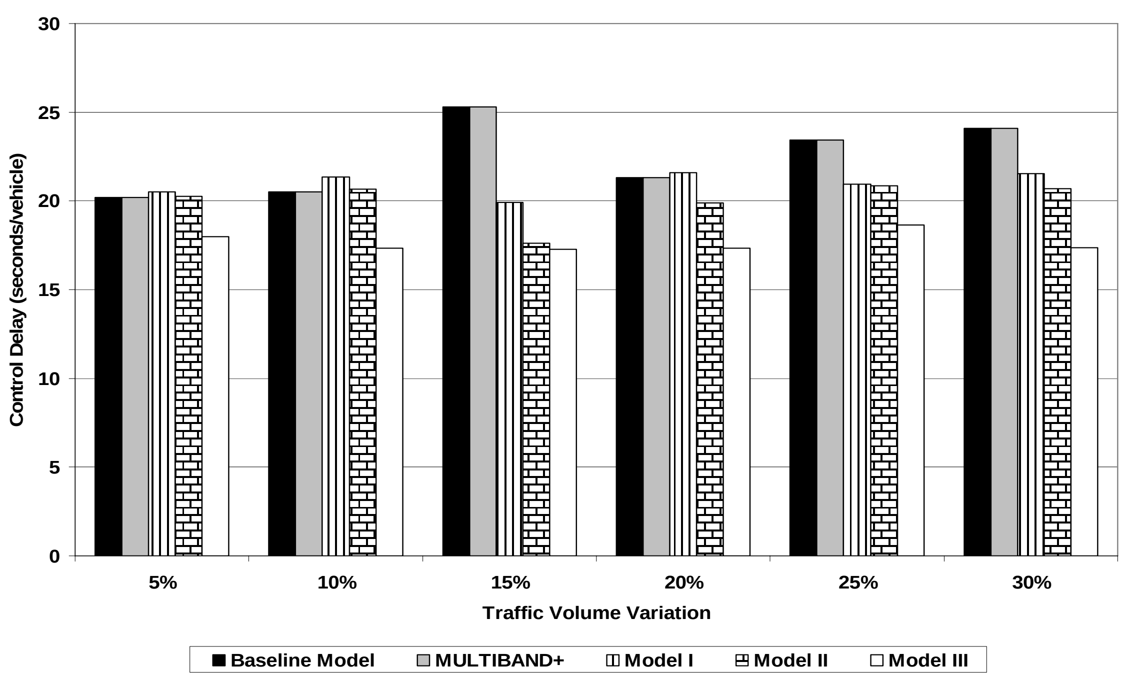

Figure 6.18 Comparison of Control Delay with Variation in Traffic Volume

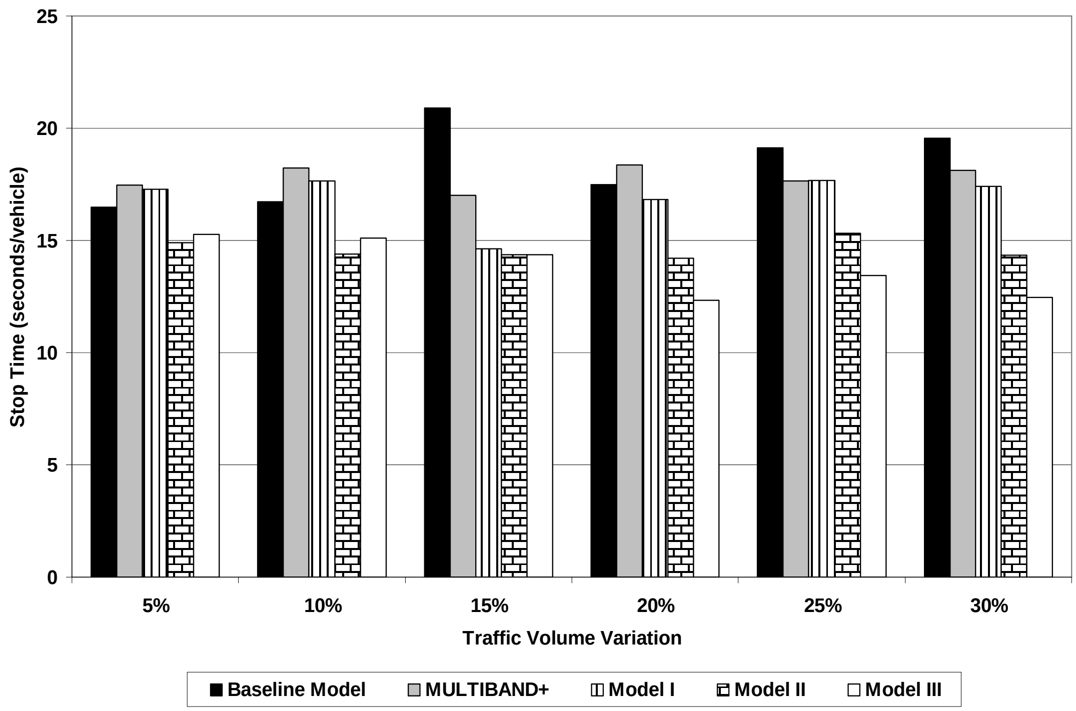

Figure 6.19 Comparison of Stop Time with Variation in Traffic Volume

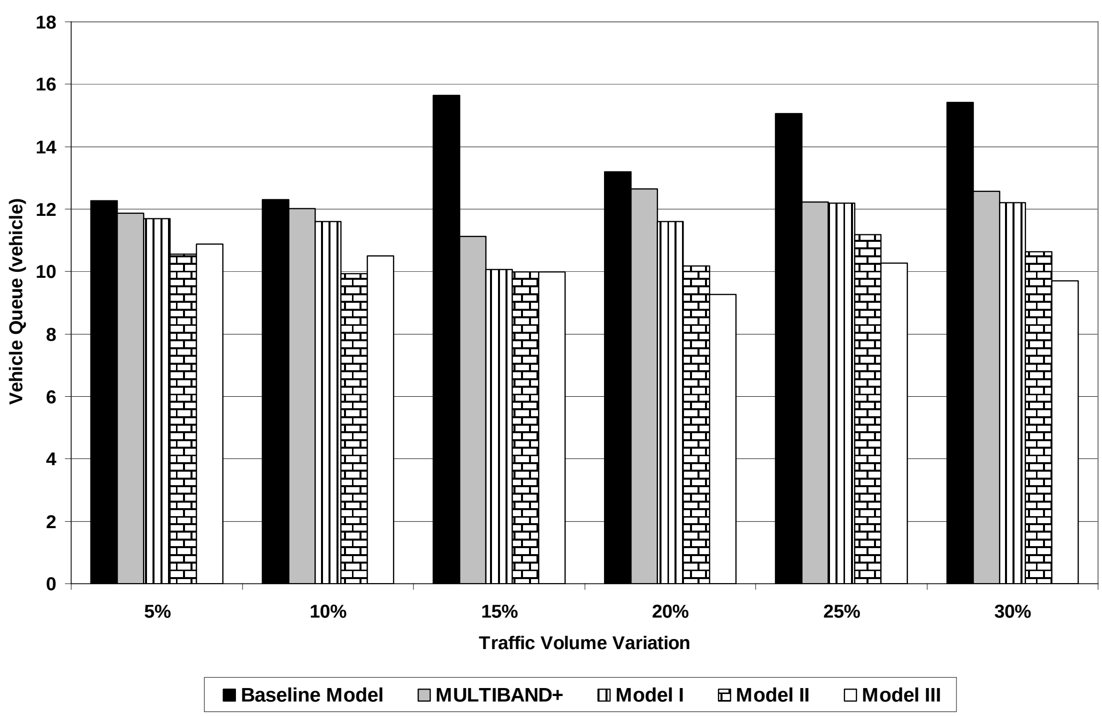

Figure 6.20 Comparison of Vehicle Queue with Variation in Traffic Volume

Table 6.10 Performance of Model IV for Different Bus Schedule Delay Scenarios for Arterial Links

Bus Schedule Delay (minutes)		Bus Delay Time					Bus Speed				
Mean	Standard Deviation	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III
5	**5**	31.14%	31.59%	30.97%	20.16%	18.76%	19.13%	21.76%	17.54%	5.37%	5.10%
10	**3**	36.29%	36.71%	36.13%	26.13%	24.83%	24.59%	27.34%	22.93%	10.20%	9.92%
15	**4**	37.87%	38.28%	37.72%	27.97%	26.70%	28.56%	31.40%	26.85%	13.71%	13.42%
20	**5**	35.91%	36.34%	35.76%	25.70%	24.39%	28.42%	31.25%	26.71%	13.58%	13.30%
25	**7**	34.96%	35.39%	34.81%	24.60%	23.27%	23.43%	26.15%	21.78%	9.17%	8.89%

Table 6.11 Performance of Model IV for Different Bus Schedule Delay Scenarios for Side Street Links

Bus Schedule Delay (minutes)		Bus Delay Time					Bus Speed				
Mean	Standard Deviation	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III
5	**5**	36.06%	39.87%	38.49%	13.22%	12.60%	24.81%	29.88%	28.20%	2.43%	1.96%
10	**3**	29.65%	33.85%	32.33%	4.53%	3.85%	33.80%	39.23%	37.44%	9.81%	9.31%
15	**4**	32.44%	36.47%	35.01%	8.31%	7.66%	25.51%	30.61%	28.92%	3.01%	2.54%
20	**5**	32.65%	36.66%	35.21%	8.59%	7.94%	33.85%	39.29%	37.49%	9.85%	9.35%
25	**7**	28.70%	32.95%	31.41%	3.23%	2.54%	34.57%	40.03%	38.23%	10.44%	9.94%

Table 6.12 Performance of Model IV for Different Bus Occupancy Scenarios for Arterial Links

Bus Occupancy		Bus Delay Time					Bus Speed				
Mean	Standard Deviation	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III
10	**3**	33.33%	33.77%	33.17%	22.71%	21.35%	24.99%	27.75%	23.33%	10.55%	10.27%
15	**5**	34.17%	34.61%	34.01%	23.68%	22.34%	25.28%	28.05%	23.61%	10.81%	10.53%
20	**7**	34.84%	35.27%	34.69%	24.46%	23.13%	21.07%	23.75%	19.46%	7.09%	6.81%
25	**3**	37.64%	38.05%	37.49%	27.70%	26.43%	23.72%	26.45%	22.07%	9.43%	9.15%
30	**5**	31.14%	31.59%	30.97%	20.16%	18.76%	19.13%	21.76%	17.54%	5.37%	5.10%

Table 6.13 Performance of Model IV for Different Bus Occupancy Scenarios for Side Street Links

Bus Occupancy		Bus Delay Time					Bus Speed				
Mean	Standard Deviation	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III	Benefits with respect to Baseline Model	Benefits with respect to MULTIBAND+	Benefits with respect to Model I	Benefits with respect to Model II	Benefits with respect to Model III
10	**3**	30.61%	34.75%	33.25%	5.83%	5.16%	33.09%	38.49%	36.71%	9.22%	8.73%
15	**5**	31.68%	35.75%	34.28%	7.27%	6.62%	34.26%	39.71%	37.91%	10.19%	9.69%
20	**7**	30.51%	34.65%	33.16%	5.69%	5.02%	34.02%	39.47%	37.67%	9.99%	9.49%
25	**3**	37.55%	41.27%	39.93%	15.24%	14.64%	32.69%	38.08%	36.30%	8.90%	8.41%
30	**5**	36.06%	39.87%	38.49%	13.22%	12.60%	24.81%	29.88%	28.20%	2.43%	1.96%

6.6 SUMMARY

This chapter presented the development of the models described in Chapters 4 and 5, and evaluated their performances through a case study, which was conducted using CORSIM applied to field data. The simulation results showed that the proposed models were superior to the Baseline Model, which used the actuated signal control implemented in the field, and the alternatives that used the MULTIBAND model for signal progression. Model II, which did not use the robust optimization technique, produced lower control delays, stop times and queue lengths when traffic volumes were relatively low. With robust optimization (Model III), benefits were significant when traffic volumes increased. Model IV, which gave priority treatment to buses, produced significant reductions in bus delays. However, Model IV did not perform as well as Model III, which did not include the bus priority functionality. Overall, Model III was the most promising. The algorithms that incorporated the models developed in this study resulted in significant reductions to control delays, stop times and queues, and increases in throughput and speeds for all scenarios.

CHAPTER 7: CONCLUSIONS AND RECOMMENDATIONS

Bus priority has emerged as one of the most promising methods for encouraging use of public transit systems. However, it is essential to recognize that a special treatment of buses may impede signal progression along an arterial and cause significant delays to automobiles. Hence, in order to develop a signal control system that not only provides bus priority but also ensures arterial progression, a robust optimization model for bus priority under arterial progression has been developed in this research.

Section 7.1 presents the conclusions, followed by recommendations for further research in Section 7.2.

7.1 CONCLUSIONS

In this study, a real-time robust optimization model for arterial signal control, which provided both signal progression and bus-priority, was developed. To solve the optimization problem, the system was decomposed into three levels, the *network* or *progression control level*, the *local* or *intersection control level*, and the *bus priority control level*.

The *progression control level* served to provide progression along the arterial. Most existing arterial progression strategies do not use real-time traffic data or use simple mathematical models to estimate evolution of traffic based on detector data. The proposed model eliminated these drawbacks by using real-time surveillance information to develop a neural network model to predict vehicle queues and arrivals. Rather than using pre-specified values, queue clearance and minimum green times were computed as functions of the predicted queues. The minimum green constraint allowed for arterial progression without unduly penalizing side streets. The bandwidth maximization problem was solved once every cycle length. To eliminate uncertainty in the prediction due to the long time horizon, robust discrete optimization technique was used to determine the progression bands. With robust discrete optimization, the progression bands that were generated were the best over all realizable traffic scenarios for the given time horizon.

At the *intersection control level*, signal timing plans were optimized for each intersection subject to bandwidth constraints to allow for uninterrupted arterial flow. Minimum green constraints were defined for safety as well as to discharge average waiting queues. Optimal signal timings were obtained every five seconds using dynamic programming approach by minimizing a performance index, which was a weighted sum of vehicle stop times, queue lengths, and intersection control delays.

At the *bus priority control level*, a bus was granted priority based on a performance index, which was a function of the schedule delay of the bus, intersection control delays and delays incurred by automobile and bus passengers, subject to bandwidth and minimum green constraints. Minimum green constraints were imposed for driver safety and to discharge average waiting queues. Bandwidth constraints allowed for uninterrupted arterial flow despite a preferential treatment of buses.

The performance of the proposed system was evaluated through a case study by applying CORSIM to field data. Six signal control alternatives were compared:

(i) Baseline Model, which used actuated signal control implemented in the field

(ii) MULTIBAND+, which used the MULTIBAND method for progression control and a short-term signal optimization model for intersection control

(iii) Model I, which used the MULTIBAND method for progression control and the proposed signal optimization model for intersection control

(iv) Model II, the proposed arterial signal control model that did not use robust optimization or bus priority functions

(v) Model III, the proposed arterial signal control model which employed robust optimization but did not give bus priority

(vi) Model IV, the proposed arterial signal control model which employed robust optimization and bus priority functions

The simulation results showed that the proposed models were superior to the Baseline Model, which used the actuated signal control implemented in the field, and the alternatives that used the MULTIBAND model for signal progression. Model II which did not use the robust optimization technique, produced lower control delays, stop times and queue lengths when traffic volumes were relatively low. With robust optimization (Model III), benefits were significant when traffic volumes increased. Model IV, which gave priority treatment to buses, produced the most reductions in bus delays. Overall, Model III was the most promising. The algorithms that incorporated the models developed in this study yielded superior results for all scenarios.

The main conclusions of this research can be summarized as follows:

1. Employed artificial neural networks to predict traffic evolution along the arterial.

To ensure signal progression along the arterial, it is necessary to reliably project traffic conditions in advance along the arterial. As indicated in Chapter 2, most existing arterial progression strategies use historical data or use simple mathematical models to estimate evolution of traffic by assuming uniform free-flow speeds. Moreover, most strategies make use of pre-specified queue clearance time. If the queue clearance time is not a function of the actual queues, the likelihood of wasted progression band due to either insufficient or excessive queue clearance time is high. This study eliminated these limitations by using real-time surveillance information to develop a neural network model to predict vehicle queues and arrivals. In addition, rather than using pre-specified values, queue clearance times were computed as functions of the predicted queues.

2. Developed an integrated system that provides in real-time a robust solution to the arterial progression problem.

In the signal progression problem, the optimal decision is dependent on the traffic demand, traffic patterns, and driver behavior. It should be noted that neither historical data nor deterministic or stochastic optimization models can fully capture uncertainties in the spatial evolution of traffic flow. As signal progression is typically provided over a long period and for multiple intersections, the discrepancy between the projected and actual arriving flows could be so significant as to mislead the optimal search process and yield ineffective progression offsets, despite using advanced prediction methods, such as artificial neural networks. Consequently, the main objective of the signal progression to provide a smooth flow of traffic along the arterial will be lost. To mitigate this problem, this study made use of robust discrete optimization, which is a comprehensive mathematical programming framework for handling uncertainties in decision-making. The system performance evaluation in Chapter 6 showed that the model that used the robust optimization technique (Model III) demonstrated the most benefits.

3. Integrated bus priority into the robust arterial signal control system.

It is evident from the review of literature that numerous bus priority strategies have been developed that range from absolute preemption techniques, with constraints only on minimum and maximum green extension times, to approaches that have integrated bus priority within signal control operations. However, most bus priority strategies are limited by their inability to handle competing requests for bus priority. Moreover, most techniques afforded priority only to those buses traveling on major roads and in the peak direction. More importantly, none of the existing bus priority systems dealt with uncertainty in traffic flow data. This study eliminated these limitations by integrating bus priority with the robust arterial signal control system developed in Chapter 4. Buses on all approaches to the intersection were considered for priority based on a performance index, which was a function of the schedule delays of all competing buses, intersection control delays and delays incurred by automobile and bus passengers. This research defined bandwidth constraints to allow for uninterrupted arterial progression and minimum green constraints to minimize excessive delays to competing traffic flows. Artificial neural networks were developed using surveillance data to predict the time taken by a bus to either join an existing queue or reach the stop bar. To contend with the ambiguities in arrival flows, bus priority was incorporated as one of the functions of the arterial signal control system that employed robust discrete optimization technique.

7.2 RECOMMENDATIONS FOR FUTURE RESEARCH

The arterial signal control models developed in this study have demonstrated to be quite promising. However, the models were evaluated in a laboratory environment using field data when available. Other than bus demand and bus route information, bus-specific data such as occupancy and schedule delays were not available. Hence, in this study bus schedule delays and occupancy were assumed to follow normal distributions. Although, sensitivity analysis was performed to validate the results, it is necessary to use real data to establish the superiority of the developed bus priority model. Some recommendations for future research are presented below:

1. Investigate robustness of the artificial neural network models developed in this research for arterial flow, traffic state and bus predictions

In this research, artificial neural network models were developed for arterial flow, traffic state and bus predictions using surveillance information. However, it should be noted that the system performance evaluation was conducted in a laboratory setting assuming the availability of all required input data. For the system to be effective in the real-world, the neural network prediction models should be able to adapt to missing or faulty data due to malfunctioning devices or broken-down communications between the field devices and the central system. Future work should investigate the robustness of the neural network models developed in this study in such instances.

Moreover, in the case study vehicle arrival detectors were placed on each lane at a location of 100 ft to 200 ft from the upstream end of the link. Sensitivity analysis should be performed to assess the impact of the location of these detectors on the accuracy of the neural network prediction models. This will help determine the need for any additional detectors if the system is to be implemented in the field.

2. Perform sensitivity analyses to assess the individual contributions of the transformation functions in determining optimal signal timing plans at the intersection control level and the bus priority control level

If the arterial signal control system is to be implemented in the real-world, one has to contend with computational efficiency issues, and potential lack of data. One way of handling this is to see if all of the transformation functions (i.e., intersection control delays, passenger delays, queue lengths, bus schedule delays, and stop times) included in this research are critical in determining the optimal signal decisions. This research assigned equal weights to the transformation functions when minimizing the objective functions at the intersection control level and the bus priority control level. A potential enhancement to this research is to conduct sensitivity analyses to determine the effects of the individual transformation functions on the optimal signal decision.

3. Field test of the robust arterial signal control model with bus priority

This research can add significant value to the areas of traffic signal control and transit priority, if the robust arterial signal control models are tested in the field and their performance evaluated in real-time. For this, one has to contend with data quality and availability issues, computational efficiency issues, and technical issues.

To circumvent the problem of erroneous or missing data due to data collection hardware and software failures, it is essential to have data filtering and imputation procedures to assess the quality of data, filter out the bad data and perform imputation to fill the missing data. Note that if the neural network models that have been enhanced to address missing data are not able to accurately predict from the available imputed data, the surveillance technology will need to be improved; otherwise, the efficacy of the proposed system will be limited.

In this research, a case study was conducted using CORSIM for an arterial of length 2.67 miles for a simulation time of 1 hour. The total execution time for the robust arterial signal control model with bus priority was 16 minutes on a laptop computer with Intel Pentium 1.8 GHz Processor with 512 MB RAM. The individual execution times for the intersection control level, the progression control level without robust optimization, and the progression control level with robust optimization, were 25%, 12.5% and 37.5%, respectively, of the total execution time. The execution time for accessing the data structures every second, which was used by all three levels, was 25% of the total execution time. For a system that also included the bus priority function, no additional computational time was needed in excess of what was required by the intersection control level. This is because the bus priority control level was called only when a bus was detected by the surveillance system; otherwise, the signal timing plans were determined by the intersection control level. To determine if the proposed system can function in real-time, it is necessary to do a hardware-in-the-loop evaluation of the system in a laboratory environment using available data prior to implementing a real-time operation of the system. Data will have to be collected for peak and off peak periods to develop the neural network prediction models. Once the hardware-in-the-loop evaluation of the system is successful, system acceptance test should be performed to demonstrate if the hardware and software components of the system can function with live real-time surveillance data.

To implement a real-time operation of the system, it is necessary to have a central system that houses the necessary hardware, such as a computer, with the robust arterial signal control logic, and a center-to-field communication. The necessary field devices are a traffic signal controller, AVL/CAD systems on each bus that provide the bus location, speed, occupancy and schedule delay, and detectors on each lane of each link to capture arriving traffic flows. Note that during the hardware-in-the-loop laboratory test if the desired accuracy of the neural network prediction models cannot be realized with the existing detectors in the field, additional detectors may have to be installed.

4. Explore potential of extending the robust arterial signal control system to support coordinated freeway, arterial and transit operations

The U.S. Department of Transportation's (USDOT) Intelligent Transportation Systems (ITS) program recently launched nine initiatives, one of which is the Integrated Corridor Management Systems (ICM) initiative (http://www.its.dot.gov/icms/index.htm). The objective of the ICM program is to reduce congestion and increase travel reliability on corridors through the coordination of freeway, arterial and transit operations. Future research should aim at investigating the benefits of integrating the robust arterial signal control system with freeway ramp meters, and variable message signs that display variable speed limits and other advisory or regulatory information.

REFERENCES

1. Balke, K. N., C. L. Dudek, and T. Urbanik II. Development and Evaluation of an Intelligent Bus Priority Concept. Presented at the 79th Annual Meeting of the Transportation Research Board, Washington, DC, 2000.
2. Bang, K. L. Optimal Control of Isolated Traffic Signals. In *Traffic Engineering and Control*, *17(7)*, 1976, pp. 288-292.
3. Bellman, R. E. *Dynamic Programming*, Princeton University Press, Princeton, 1957.
4. Bowen, G. T., R. D. Bretherton, J. R. Landles, and D. J. Cook. Active Bus Priority in SCOOT. Proceedings of the Seventh International Conference on *Road Traffic Monitoring and Control*, London, England, IEE Conference Publication No. 391, 1994, pp. 73-76.
5. Chang, E. C., and C. J. Messer. Minimum Delay Optimization of a Maximum Bandwidth Solution to Arterial Signal Timing. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, DC, 1985, pp. 89-95.
6. Chang, E. C., C. J. Messer, and B. G. Marsden. Reduced-Delay Optimization and other Enhancements in the PASSER II-84 Program. In *Transportation Research Record 1005*, TRB, National Research Council, Washington, DC, 1985, pp. 80-89.
7. Chang, G. L., M. Vasudevan, and C.C. Su. Bus-Preemption under Adaptive Signal Control Environments. In *Transportation Research Record 1494*, TRB, National Research Council, Washington, DC, 1995, pp. 146-154.
8. Chang, G. L., and C. C. Su. Predicting Intersection Queue with Neural Network Models. In *Transportation Research*, *3C(3)*, 1995, pp. 175-191.

9. Cheu, R. L. and S. G. Ritchie. Automated Detection of Lane-Blocking Freeway Incidents Using Artificial Neural Networks. In *Transportation Research*, vol. 3C, No. 6, 1996, pp. 371-388.
10. Cheu, R. L., D. Srinivasan, and W. H. Loo. Training Neural Networks to Detect Freeway Incidents using Particle Swarm Optimization. Presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC, 2004.
11. Chien, S., X. Liu, and K. Ozbay. Predicting Travel Times for South Jersey Real-Time Motorist Information System. Presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC, 2004.
12. Cho, H., and L. R. Rilett. Forecasting Train Travel Time. Presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, DC, 2003.
13. Cohen, S. L. Concurrent Use of MAXBAND and TRANSYT Signal Timing Programs for Arterial Signal Optimization. In *Transportation Research Record 906*, TRB, National Research Council, Washington, DC, 1983, pp. 81-84.
14. Cohen, S. L., and C. C. Liu. The Bandwidth-Constrained TRANSYT Signal Optimization Program. In *Transportation Research Record 1057*, TRB, National Research Council, Washington, DC, 1986, pp. 1-9.
15. Conrad, M., F. Dion, and S. Yagar. Real-Time Traffic Signal Optimization with Transit Priority: Recent Advances in the SPPORT Model. Presented at the 77th Annual Meeting of the Transportation Research Board, Washington, DC, 1998.
16. Cornwell, P. R. Dynamic signal co-ordination and public transport priority. *Road Traffic Monitoring and Control*, 1986, pp. 158-161.

17. *CORSIM User's Guide*, U.S. Department of Transportation, Federal Highway Administration, Washington, DC, 1999.

18. Courage, K. C., C. E. Wallace, and J. A. Wattleworth. Effect of Bus Priority System Operation on Performance of Traffic Signal Control Equipment on NW 7th Avenue. Prepared for the U.S. Department of Transportation, Report No. UTMA FL-06-0006, 1977.

19. Dell 'Olmo, P., and P. Mirchandani. REALBAND: An Approach for Real-Time Coordination of Traffic Flows on Networks. In *Transportation Research Record 1494*, TRB, National Research Council, Washington, DC, 1995, pp. 106-116.

20. Dougherty, M., and M. R. Cobbett. Short-Term Inter-Urban Traffic Forecasts Using Neural Networks. In *International Journal of Forecasting*, No. 31, 1997, pp. 21-31.

21. Duerr. P. A. Dynamic Right-of-Way for Transit Vehicles: An Integrated Modeling Approach for Optimizing Signal Control on Mixed Traffic Arterials. Presented at the 79th Annual Meeting of the Transportation Research Board, Washington, DC, 2000.

22. Faussett, L. Fundamentals of Neural Networks - Architectures, Algorithms and Applications, Prentice-Hall, 1994.

23. Garrow, M., and Machemehl, R. Development and Evaluation of Transit Signal Priority Strategies. Presented at the 77th Annual Meeting of the Transportation Research Board, Washington, DC, 1998.

24. Gartner, N. H. OPAC: A Demand-Responsive Strategy for Traffic Signal Control. In *Transportation Research Record 906*, TRB, National Research Council, Washington, DC, 1983, pp. 75-81.

25. Gartner, N. H., S. F. Assmann, F. Lasaga, and D. L. Hou. MULTIBAND - A Variable-Bandwidth Arterial Progression Scheme. In *Transportation Research Record 1287*, TRB, National Research Council, Washington, DC, 1990, pp. 212-222.

26. Gartner, N. H., P. J. Tarnoff, and C. M. Andrews. Evaluation of Optimized Policies for Adaptive Control Strategy. In *Transportation Research Record 1324*, TRB, National Research Council, Washington, DC, 1991, pp. 105-114.

27. Grubba, J., D. Allyn, J. Barbaresso, and R. Beaubien. Selecting Real-Time Adaptive Traffic Control for Oakland County, Michigan. Presented at the 72nd Annual Meeting, Transportation Research Board, Washington, DC, 1993.

28. Haykin, S. Neural Networks - A Comprehensive Foundation, Macmillan, 1994.

29. Head, K. L., P. B. Mirchandani, and D. Sheppard. Hierarchical Framework for Real Time Traffic Control. In *Transportation Research Record 1360*, TRB, National Research Council, Washington, DC, 1992, pp. 82-88.

30. Head, K. L., P. B. Mirchandani, and S. Shelby. The RHODES Prototype: A Description and Some Results. Presented at the 77th Annual Meeting of the Transportation Research Board, Washington, DC, 1998.

31. Henry, J. J., J. L. Farges, and J. Tuffal. The PRODYN Real Time Traffic Algorithm. *IFAC Symposium on Control in Transportation Systems*, 1983, pp. 305-310.

32. Heydecker, B. G. Capacity at a signal-controlled junction where there is priority for buses. In *Transportation Research*, *17B(5)*, 1983, pp. 341-357.

33. *Highway Capacity Manual*, Special Report 209, Third Edition, Transportation Research Board, Washington, DC, 1997.

34. Hounsell, N., J. Landles, R. D. Bretherton, and K. Gardner. Intelligent Bus Priority in London: Evaluation and Exploitation in INCOME. 6th World Congress on Intelligent Transport Systems, Toronto, Canada, 1999.

35. Huang, S. H., and B. Ran. An Application of Neural Network on Traffic Speed Prediction Under Adverse Weather Condition. Presented at the 82nd Annual Meeting of the Transportation Research Board, Washington, DC, 2003.

36. Jacobson, J., and Y. Sheffi. Analytical model of traffic delays under bus signal preemption: theory and application. In *Transportation Research*, *15B(2)*, 1981, pp. 127-138.

37. Jeong, R., and L. R. Rilett. The Prediction of Bus Arrival Time using AVL data. Presented at the 83rd Annual Meeting of the Transportation Research Board, Washington, DC, 2004.

38. Jin, X., R. L. Cheu, and D. Srinivasan. Development and Adaptation of Constructive Probabilistic Neural Network for Freeway Incident Detection. In *Transportation Research*, vol. 10C, No. 2, 2002, 121-147.

39. Khasnabis, S., G. V. Reddy, and B. B. Chaudry. Signal preemption as a priority treatment tool for transit demand management. *Vehicle Navigation and Information System Conference Proceedings*, Paper No. 912865, Dearborn, Michigan, 1991.

40. Kouvelis, P., and G. Yu. Robust Discrete Optimization and Its Applications, Kluwer Academic Publishers, Norwell, MA, 1997.

41. Lan, C. C., C. J. Messer, N. A. Chaudhary, and E. C. Chang. A Compromised Approach to Optimize Traffic Signal Coordination Problems During Unsaturated

Conditions. Presented at the 71st Annual Meeting of the Transportation Research Board, Washington, DC, 1992.

42. Little, J., M. D. Kelson, and N. H. Gartner. MAXBAND: A Program for Setting Signals on Arteries and Triangular Networks. In *Transportation Research Record 795*, TRB, National Research Council, Washington, DC, 1981, pp. 40-46.
43. Lowrie, P. R. The Sydney Coordinated Adaptive Traffic System: Principles, Methodology, Algorithms. Proceedings of the International Conference on *Road Traffic Signaling*, London, England, IEEE Conference Publication No. 207, 1982, pp. 67-70.
44. Ludwick, J. S. Bus Priority System: Simulation and Analysis, Final Report, The Mitre Corporation. Prepared for the U.S. Department of Transportation, Report No. UTMA-VA-06-0026-1, 1976.
45. Luk, J. Y. K., A. G. Sims, and P. R. Lowrie. SCATS: Application and Field Comparison with a TRANSYT Optimized Fixed Time System. Proceedings of the International Conference on *Road Traffic Signaling*, London, England, IEE Conference Publication No. 207, 1982, pp. 71-74.
46. MacGowan, C. J., and I. J. Fullerton. Development and testing of advanced control strategies in the urban traffic control system. *Public Roads*, *43(3)*, 1979, pp. 97-105.
47. Mark, C. D., A. W. Sadek, and D. Rizzo. Predicting Experienced Travel Time with Neural Networks: A Paramics Simulation Study. Presented at the IEEE Intelligent Transportation Systems Conference, Washington, DC, 2004.
48. Martin, P. T. and S. Hockaday. SCOOT - An Update. In *ITE Journal 65 (1)*, January 1995, pp. 44-48.

49. Mauro, V., and C. Di. Taranto. UTOPIA. Proceedings of the 6th *IFAC/IFIP/IFORS Symposium on Control, Computers, and Communication in Transportation*, Paris, France, 1989, pp. 245-252.

50. McCulloch, W. S., and W. A. Pitts. A Logical Calculus of the Ideas Imminent in Nervous Activity, *Bulletin of Mathematical Biophysics*, vol. 5, pp 115-133, 1943.

51. McGinley, F. J., and D. R. Stolz. The design of tram priority at traffic signals. *Journal of Advanced Transportation 19:2i*, 1985, pp. 133-151.

52. Messer, C. J., R. H. Whitson, C. L. Dudek, and E. J. Romano. A Variable Sequence Multiphase Progression Optimization Program. In *Highway Research Record 445*, Highway Research Board, National Research Council, Washington, DC, 1973, pp. 24-33.

53. Miller, A. J. A Computer Control System for Traffic Network. Proceedings of the Second International Symposium on the *Theory of Road Traffic Flow*, Paris, France, 1963, pp. 200-220.

54. MPL Modeling System, 4.2, Maximal Software, Inc., 2004.

55. Owen, L. E., C. M. Stallard, and Y. Zhang. Evaluation of the Reston Parkway RT-TRACS Field Test, ITT Industries, Inc., Systems Division. Prepared for the U.S. Department of Transportation, Federal Highway Administration, March 31, 1999.

56. *PASSER II-90 User's Guide*, Series of Methodology for Optimizing Signal Timing (MOST), Volume 3, U.S. Department of Transportation, Federal Highway Administration, Washington, DC, 1991.

57. Pooran, F. J., C. Andrews, and N. H. Gartner. Implementation of the OPAC Adaptive Control System in RT-TRACS. Presented at the *ITS World Congress Sixth Annual Meeting*, Toronto, Canada, November 1999.

58. QNET 2000, Neural Network Modeling for Windows, Vesta Services, Inc., 1999.

59. Richardson, A. J., and K. W. Ogden. Evaluation of active bus-priority signals. In *Transportation Research Record 718*, TRB, National Research Council, Washington, DC, 1979, pp. 5-12.

60. Salter, R. J., and J. Shahi. Prediction of effects of bus-priority schemes by using computer simulation techniques. In *Transportation Research Record 718*, TRB, National Research Council, Washington, DC, 1979, pp. 1-5.

61. Sen, S., and K. L. Head. Controlled Optimization of Phases (COP) at an Intersection. In *Transportation Science*, *Vol. 31 (1)*, 1997, pp. 5-17.

62. Seward, S. R., and R. N. Taube. Methodology for evaluating bus-actuated, signal-preemption systems. In *Transportation Research Record 630*, TRB, National Research Council, Washington, DC, 1977, pp. 11-17.

63. Skabardonis, A. Control Strategies for Transit Priority. Presented at the 79th Annual Meeting of the Transportation Research Board, Washington, DC, 2000.

64. Taylor, W. C., and A. R. Ahmed. Analysis of Corridor Delay under SCATS Control - Orchard Lake Road, Final Report. Submitted to Oakland County Road Commission, 1997.

65. Vincent, R. A., B. R. Cooper, and K. Wood. Bus-Actuated Signal Control at Isolated Intersections-Simulation Studies of Bus Priority. *Transport and Road Research Laboratory Report 814*, Crowthorne, U.K., 1978.

66. Vincent, R. A., and J. R. Peirce. MOVA: Traffic Responsive, Self-Optimizing Signal Control for Isolated Intersections. *Transport and Road Research Laboratory Report 170*, Crowthorne, U.K., 1988.

67. Wallace, C. E., and K. G. Courage. Arterial Progression – New Design Approach. In *Transportation Research Record 881*, TRB, National Research Council, Washington, DC, 1982, pp. 53-59.

68. Wallace, C. E. TRANSYT-7F User's Manual, Report UF-TRC-U32 FP-06/07. U.S. Department of Transportation, Federal Highway Administration, 1984.

69. Yagar, S. Efficient Transit Priority at Intersections. In *Transportation Research Record 1390*, TRB, National Research Council, Washington, DC, 1993, pp. 10-15.

70. Yagar, S. and B. Han A Procedure for Real-Time Signal Control that Considers Transit Interference. In *Transportation Research*, *28B(4)*, 1994, pp. 315-331.