SIGNAL TIMING OPTIMIZATION FOR RELIABLE AND SUSTAINABLE MOBILITY

By

LIHUI ZHANG

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2010

To my family

ACKNOWLEDGMENTS

First and foremost, I would like to express my gratitude to my supervisor, Dr. Yafeng Yin, for his guidance through my PhD study. Without his support and encouragement, this dissertation would never have been finished.

I am grateful to the transportation research center for providing me multiple resources for transportation study, including the wonderful courses, seminars, distinguished lectures and student conferences, which will be of great help to my future development, both academically and professionally.

I would like to thank Drs. Siriphong Lawphongpanich, Lily Elefteriadou, Scott Washburn and Sivaramakrishnan Srinivasan for serving as my committee members, and for their valuable comments and suggestions on this dissertation.

I would also like to thank my colleagues in the transportation research center, especially Yingyan Lou, Ziqi Song, Di Wu for the discussions on a broad range of topics, and for their feedbacks to my research work.

Finally, I want to thank my family for their love, and their support on any of the decisions I made. Special thanks go to my wife, Xiyuan Li, for her company in this foreign country, and for the happiness and joy she brings to my life.

TABLE OF CONTENTS

page

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

SIGNAL TIMING OPTIMIZATION FOR RELIABLE AND SUSTAINABLE MOBILITY

By

Lihui Zhang

August 2010

Chair: Yafeng Yin
Major: Civil Engineering

This dissertation develops a stochastic programming approach to proactively consider a variety of uncertainties associated with signal timing optimization for fixed-time or actuated traffic signals. Representing the uncertain parameter of interest as a number of scenarios and the corresponding probabilities of occurrence, the stochastic programming approach optimizes signal timings with respect to a set of high-consequence or worst-case scenarios. The resulting signal timing plans produce smaller delays and less vehicular emissions under those scenarios, thereby leading to more reliable and sustainable mobility.

To illustrate the stochastic programming approach, below are three applications in traffic signal timing. The first application is to optimize the settings of fixed-time signals along arterials under day-to-day demand variations or uncertain future traffic growth. Based on a cell-transmission representation of traffic dynamics, an integrated stochastic programming model is formulated to determine cycle length, green splits, phase sequences and offsets that minimize the expected delay incurred by high-consequence scenarios of traffic demand. The stochastic programming model is simple in structure but contains a large number of binary variables. Existing algorithms, such as branch and bound, are not able to solve it efficiently. Consequently, a simulation-based genetic algorithm is developed to solve the model. The model and algorithm

are validated and verified using two networks, under congested and uncongested traffic conditions.

The second application further considers traffic emissions, and develops a bi-objective optimization model to make an explicit tradeoff between traffic delays and roadside human emission exposure. Based on the cell-transmission representation of traffic dynamics, a modal sensitive emission approach is used to estimate the tailpipe emission rate for each cell of a signalized arterial. A cell-based Gaussian plume air dispersion model is then employed to capture the dispersion of air pollutants and compute the roadside pollutant concentrations. Given a stochastic distribution of the wind speed and direction of a corridor, a scenario-based stochastic program is formulated to optimize the cycle length, phase splits, offsets and phase sequences of signals along a corridor simultaneously. A genetic algorithm is further developed to solve the bi-objective optimization problem for a set of Pareto optimal solutions. The solutions form an efficient frontier that presents explicit tradeoffs between total delay of the corridor and the human emission exposure of the roadside area incurred by high-consequence scenarios.

The last application is to synchronize actuated signals along arterials for smooth and stable progression under uncertain traffic conditions, mainly addressing the issue of uncertain (not fixed) starts/ends of green of sync phases. The model developed is based on Little's mixed-integer linear programming (MIP) formulation, which determines, e.g., offsets and progression speed adjustment, to maximize the two-way bandwidth. By specifying scenarios as realizations of uncertain red times of sync phases, the regret associated with a coordination plan is defined with respect to each scenario, and then a robust counterpart of Little's model is formulated as another MIP to minimize the average regret incurred by a set of high-consequence scenarios. The numerical example shows that the resulting robust coordination plan is able to increase the

worst-case and 90th percentile bandwidths by approximately 20% without compromising the average bandwidth.

In summary, these three applications demonstrate that the proposed stochastic programming approach is valid for signal timing optimization under uncertainty. The resulting timing plans are expected to perform more robustly and effectively in uncertain environments, thereby making the transportation system more reliable and sustainable.

CHAPTER 1
INTRODUCTION

1.1 Background

Congestion has been a severe problem in many metropolitan areas, and is getting worse. While the largest cities are the most congested, congestion occurs in regions of all sizes. Peak periods are becoming longer, stop-and-go traffic conditions can last the entire day in some cities. From 1993 to 2003, travelers in the U.S. spent an extra seven hours per year in travel time; the percentage of congested freeway mileage increased from 51% to 60%, although the total mileage was increasing; trips in peak period averagely experienced 37% more time than a free-flow trip, a 9% increase. In 2007, congestion cost reached $87.2 billion, an increase of more than 50% over the previous decade (Schrank and Lomax, 2009).

Traffic control devices are recognized as one of the major causes of traffic congestion. Disruption of traffic flow by control devices contributes much to traffic congestion, and improperly timed traffic signals account for 5% of the total congestion (Schrank and Lomax, 2009). A recent survey (NTOC, 2007) on the quality of traffic signal operations in the United States concluded that the nation scored a “D” in terms of the overall quality of traffic signal operations. In another study, ORNL (2004) suggests that 56% of the traffic signals in U.S. are in urgent need of re-optimization.

Given a large segment of signal control systems in use today are still fixed-time or actuated, further improvements in their efficiency (e.g., delay per vehicle) and robustness (e.g., variance of delay per vehicle) of signal control systems can yield significant improvement in the management of traffic flows and level of congestion. Studies around the country have shown that improved signal timing typically requires little or no infrastructure costs and produced a benefit to cost ratio of 40 (NTOC, 2007). The benefits include shorter commute times, improved air

quality, reduction in certain types and severity of crashes, and reduced driver frustration (FHWA, 2008). According to the 2009 Urban Mobility Report (Schrank and Lomax, 2009), signal coordination in 439 urban areas relieved a total of 19.6 million hours of delay in 2007, a value of $404 million.

In the current practice, fixed-time or actuated signal systems typically segment a day into a number of time intervals, each of which is assigned a best suited signal timing plan as determined by applying Webster's formula or using optimization tools such as TRANSYT-7F. Typically, three to five signal timing plans are used in a given day. For such a system to work well, the traffic pattern within each interval should remain relatively constant. Unfortunately, travel demands and traffic arrivals at intersections in practice can vary significantly even for the same time of day and day of week. The signal timings obtained from traditional approaches are thus generally unstable and unreliable under such traffic conditions.

On the other hand, because of its negative effects on health and living conditions, air pollution has long emerged as one of the most acute problems in many metropolitan areas. A major source of this pollution is the emissions from vehicular traffic. They contribute considerably to the level of carbon monoxide (CO), nitrogen oxides (NOx), and volatile organic compounds (VOCs) in the environment. They are also responsible for approximately 20% of greenhouse gas emissions (Poole, 2009). In order to achieve a more sustainable mobility, reducing traffic emissions has been an ongoing endeavor of many governmental authorities over the past two decades. However, many emission reducing programs implemented thus far are passive in nature because they regulate only the rates (e.g., grams per mile) at which a vehicle generates these hazardous air pollutants. Depending on the amount of driving or the mode of driving, cars meeting an emission standard can generate as much air pollution as cars not

meeting the standard. More proactive strategies that aim to reduce the mileage traveled by vehicles or improve the efficiency of the transportation network have recently gained more and more attention (e.g., Cambridge Systematics, 2009). Signal timing imposes a huge impact on traffic emissions because it interrupts traffic flow (for good reasons) and creates additional deceleration, idle and acceleration driving modes to the otherwise cruise driving mode. Traffic emissions are very sensitive to driving modes. For example, Coelho et al. (2005) reports that decelerating vehicles produce 125% more CO than those in the cruise driving mode, and Rakha et al. (2000) showed that an efficient signal coordination can reduce emissions up to 50%, in a highly simplified scenario.

This dissertation proposes a stochastic programming approach to design signal timing under uncertain traffic conditions and determine more sustainable signal timing to reduce harmful traffic emissions.

1.2 Dissertation Outline

The remainder of the dissertation is organized as follows. Chapter 2 reviews the literature of signal timing optimization under different network configurations. Chapter 3 introduces the uncertainties associated with traffic modeling, and discusses the stochastic programming approach for optimizing signal timings. The next three chapters apply the approach to three different signal timing problems in practice: Chapter 4 optimizes the timings of fixed-time signals under day-to-day demand variations, Chapter 5 considers the tradeoff between delays and roadside human emission exposure to generate more sustainable signal timing plans, and Chapter 6 synchronizes actuated signals on arterials, addressing the issue of uncertain starts/ends of sync phases. Finally, conclusions and future work are given in Chapter 7.

CHAPTER 2
LITERATURE REVIEW

Traffic signals have evolved considerably since they were first introduced to prevent collision in London in 1868. The first automatic traffic signal was implemented in Detroit, while the first vehicle-actuated traffic signal was installed in Baltimore in 1928 (Hensher et al., 2001). Signal timing offers the opportunity to improve the mobility of a transportation system and also prevent the environmental deterioration.

2.1 Traffic Signal Operation

Traffic signals are generally operated in three modes: fixed-time, actuated and adaptive control.

2.1.1 Fixed-Time Control

Fixed-time control is the simplest, less expensive and easier to maintain. The signals assign right-of-way at intersections according to predetermined schedules, i.e., timing plans. The phase sequence, phases splits, cycle length and (or) offset for each signal are fixed, and determined based on historical traffic pattern. Because it does not account for any traffic demand variations, fixed-time signal control may cause additional delay (FHWA, 2008; Boillot et al., 1992).

2.1.2 Actuated Control

Signal timing in actuated control, in contrast, consists of intervals that are called and extended in response to vehicle activations. The traffic controller attempts to adjust green time continuously and in some cases, the sequence of phasing. These adjustments occur in accordance with real-time measures of traffic demand from vehicle detectors placed at the intersection approaches. Depending on the settings of the controller, the adjustments are constrained by necessary controller parameters. Actuated control usually reduces delay, increases capacity and

can be safer than the fixed-time control, though they are more expensive to implement and also require advanced training of practitioners to operate properly (FHWA, 2008; Boillot et al., 1992).

Traffic-actuated control can be of two types, semi-actuated and fully actuated control, depending on the traffic approaches to be detected.

In semi-actuated control, the monitored phases include any protected left-turn phases and phases of the side streets. The major movements are called “sync” phase, served unless there is a conflicting call on a minor movement phase. Minor movement phases receive green only after the sync-phase yield point and are terminated on or before their respective force-off points. These points occur at the same point during the background signal cycle and ensure that the major road phase will be coordinated with the adjacent signals. If there are no calls present at the yield point, the non-coordinated phases will be skipped for an entire cycle length. The major disadvantage of semi-actuated control is that continuous demand on the minor phases can cause excessive delay to the major movements if the maximum green and passage time parameters are not set appropriately (FHWA, 2008).

In fully-actuated control, vehicle detectors are installed on all traffic approaches. For each phase, there is a set of minimum and maximum green time. If there are no opposing vehicles that waiting for the right-of-way, the moving traffic will receive additional green time. Fully actuated signals are mostly found at intersections that exhibit large fluctuations of traffic volumes from all of the approaches during the day.

Much of the benefit of traffic-actuated control is derived from the ability of the controller’s proactively responding to the fluctuations in traffic volume, which provides greater efficiency compared to fixed-time control by servicing cross-street traffic only when required. The primary disadvantage of fixed-time control is avoided as the main street traffic is not interrupted

unnecessarily (Jongenotter and Monsma, 2002). This is particularly beneficial during off-peak conditions, resulting in fewer stops and smaller delays to the traffic on the major arterial, which ultimately leads to a decrease in fuel consumption and pollutant emissions. However, actuated traffic signal can only respond to the traffic flow fluctuation to a certain degree. A retiming is needed after a period of time to ensure its efficiency.

2.1.3 Adaptive Control

Adaptive systems control traffic signals across an arterial or a large network systematically. These kinds of systems rely on advanced detection and information technologies and increasing computation speed to adjust the lengths of signal phases based on solving certain optimization problems in every few seconds. With such a mechanism, adaptive signal systems are obviously more capable of optimizing signal timings against fluctuating traffic conditions, and generally can save up to 10% in total travel time (Boillot, 1992). On the other hand, these systems are expensive, beyond the budget of many agencies.

2.2 Signal Timing Optimization

Since the introduction of automatic traffic signals, much work has been done to develop methodologies for timing signals. The configuration of the transportation network under consideration can have a significant impact on how its traffic signals are timed.

2.2.1 Isolated Intersection

Isolated traffic signals can be timed without considering other adjacent signals, allowing the flexibility of setting timings that optimize different objectives for individual intersections.

During the early stage, methods to determine the fixed signal timings were developed assuming the traffic arrival from every intersection approach has a constant headway, which is not realistic in the field (Matson et al., 1955). Webster (1958) recognized the uncertain nature of

the traffic arrivals and developed the following delay equation to estimate the average delay of vehicles:

$$d = \frac{c(1-\lambda)^2}{2(1-\lambda x)} + \frac{x^2}{2q(1-x)} - 0.65\left(\frac{c}{q^2}\right)^{1/3} x^{(2+5\lambda)}$$

where, d is the average delay per vehicle; c is the cycle length; λ is the effective green in proportion of cycle length; q is the flow rate; s is the saturation flow rate; x is the degree of saturation, equal to $q/(\lambda s)$. The second term in the equation captures the random nature of the traffic arrivals, for the average delay experienced by a Poisson stream of traffic.

By further assuming that the effective green times of the phases are proportional to their respective degree of saturation, Webster (1958) determined the cycle length using the following equation:

$$C_o = \frac{1.5\sum tL_i + 5}{1.0 - \sum x_i}$$

where, C_o is the optimal cycle length; tL_i is total lost times per cycle; x_i is the degree of saturation for phase i.

Note that the equation is valid only when the sum of x_i is less than 1.0, and all the approaches within the same phase will need to have the same degree of saturation, which are both easy to be violated under congested traffic conditions.

Allsop (1971, 1972) presented a more rigorously complete treatment of the isolated signal optimization problem, still using Webster's delay formula, and explored the delay with respect to the traffic volume used to calculate signal timings and concluded that the delay errors are especially serious when several volumes are measured incorrectly simultaneously.

On the other hand, research was carried out to investigate signal timing under over-saturated traffic conditions. Gazis (1964) first considered over-saturated traffic condition and

optimized signals in a dynamic setting. His approach first allocates the maximum green to the major road and the minimum green to the minor road, then switches the green time between these two to balance the residual queues. The TRANSYT method (Robertson, 1969) is similar to Gazis's idea, and requires a detailed input of network characteristics and traffic flows. The central element of the TRANSYT model is flow profiles. Within a user specified modeling horizon, the model estimates the performance index as a combination of travel time and stops, and then gradually adjusts the current signal plan until no improvement of the performance index can be made.

Michalopoulos and Stephanopoulos (1977) modified Gazis' approach and optimized single intersections with state variable constraints. The optimal policy minimizes intersection delay subject to queue length constraints, and switches the signals as soon as the queues reach their limits to balance the input and output flows. Chang and Lin (2000) developed a discrete state space version of Gazis's model.

Lan (2004) proposed a new optimal cycle length methodology for near-saturated or over-saturated conditions, where Webster's equation fails. A nonlinear programming problem was formulated with the delay equation adopted from the Highway Capacity Manual (TRB, 2000). The optimal simple cycle length formulation was fitted using field samples, which shows very small deviation from the analytical solution.

Miller (1965) explored actuated control at an isolated intersection. The discrete model developed can determine the best time to switch phase, taking into account the lost time due to changing phases. The system was demonstrated in a simulation study to reduce delay by a maximum of 40% compared with the standard vehicle actuated case. Many other studies also focus on actuated controllers. Head (2006) represented the actuated control logic with a

procedure graph, which can act as basis to support decision making like transit priority control. Liu and Bhimireddy (2009) developed an analytical model for vehicle-actuated signal at an isolated intersection to study the effects of "one time green extension for vehicle platoon on major approach" on the delays in subsequent cycles. The model extends the queuing model developed by Mirchandani and Zou (2007) to account for vehicle actuations. Comert et al. (2009) incorporated the queue length into actuated signal logic for an isolated intersection to determine the maximum green in each cycle. The resulting timing plans were demonstrated to be efficient under fluctuated traffic conditions. Agbolosu-Amison and Park (2009) optimized timing plans under dynamic gap-out feature using a stochastic optimization method and software-in-the-loop simulation. The resulting timing plan was demonstrated using VISSIM to reduce vehicular delays by 12.5% compared with regular gap-out feature. Viti and Zuylen (2009) presented a modeling approach for actuated signal timings based on the probabilistic theory, allowing the computation of variables like vehicle waiting time with any arrival distribution. HCM 2000 also provides a reasonable approximation of the operation of actuated controllers for nearly all the conditions encountered in practice, and the results have correlated well with extensive simulation data, though the procedure assumes constant departure headway.

The above studies have a central assumption that the traffic flows at the intersection are given as constant average flow rate for each approach, which appears to be a great barrier to the effective design of traffic signal timings to achieve reasonable targets (Smith et al., 2001).

Heydecker (1987) pointed out that once signal settings have been calculated, they would normally be used in a range of circumstances. For example, a single signal plan may be used throughout the morning peak period on each day of a week. If the parameters used in the calculation of the signal settings vary, then a number of realizations will be observed. If signal

settings are calculated to optimize some measure of signal performance using the mean values of these observations, then they will not be optimal for the observations themselves, because 1) most measures of signal performance vary in a nonlinear way with respect to some of the parameters; 2) the correlations between variations in the parameter values may give rise to further nonlinearity in the measure of junction performance. Heydecker investigated the consequences of variability in traffic flows and saturation flows for the calculation of signal settings and then proposed an optimization formulation that minimizes the mean rate of delay over the observed arrivals and saturation flows. Following the same notion, Ribeiro (1994) proposed a novel technique called Grouped Network, using TRANSYT to calculate timing plans that are efficient even when demand is variable. Both studies focus on optimizing the average performance. Yin (2008b) developed three signal optimization models: scenario-based mean-variance optimization, scenario-based conditional value-at-risk minimization, and min-max optimization to determine robust signal timing for isolated signals, which performs better against worst case scenarios.

Much recent research has focused on adaptive signal controls. The work on single intersection can be traced back to Newell (1998), who proposed an adaptive traffic control strategy based on the queuing model. Mirchandani and Zou (2007) developed an approach to evaluate this adaptive system. Li and Prevedouros (2004) used a hybrid optimization and rule-based strategy to control an isolated oversaturated intersection. The objective of the model is to maximize the throughput and control queue length. MOVA (1988) used a similar method for isolated signal control. In recognition of their capability in providing up-to-date signal timings, these adaptive models are able to better tolerate fluctuating traffic arrivals.

2.2.2 Arterial

For intersections located along a major arterial, isolated operations can be improved by considering coordination of the major movements along the arterial. Common cycle lengths are often employed to facilitate this coordination. For most arterial streets with signal spacing between 500 feet and 0.5 mile, coordinated operation can often yield benefits by improving progression between signals. On arterials with higher speeds, it can be beneficial to coordinate signals spaced a mile apart or even longer. Signals that are located very close together often require settings that manage queues rather than progression as the dominant policy (FHWA, 2008).

Morgan and Little (1964) first considered to maximize the bandwidth of "green wave" to increase the chance of vehicles' traversing an arterial without stop. Little's MILP formulation (Little, 1966) maximizes the two-way bandwidth to synchronize signals along arterials by determining offsets and progression speed adjustment etc. The model has been proved to be a flexible and robust approach for signal synchronization and actually lays the foundation for MAXBAND. Gartner and Stamatiadis (2002) followed the similar concept and formulated a multiband/multiweight optimization program for arterials, which can potentially tolerate different traffic patterns. Rather than considering the green band, Allsop(1968) optimized offsets by minimizing the total system delay.

The above studies deal with fixed-time signal systems, while for actuated systems, the problem becomes more complex due to the fact that the starts of green of the sync phases (typically Phases 2 and 6) are not fixed. Several approaches have been proposed in the literature to address such a so-called "early return to green" problem in determining the offsets.

Jovanis and Gregor (1986) suggested adjusting the end of green of the sync phases to the end of the through-band for non-critical signals. Skabardonis (1996) proposed three methods for

determining offsets for actuated signals from the optimal fixed-time splits and offsets. Although the three methods differ in the procedure and applicable situation, the concepts are essentially the same: making best estimates of the average starting point of the sync phases and then optimizing the offset based on the estimates. Chang (1996) offered a similar suggestion of obtaining the offsets from a second optimization run that uses the anticipated green times of the non-coordinated phases as constraints on their maximum green times. Park et al. (2000) formulated an optimization problem to determine the cycle length, offsets, green split and phase sequence simultaneously. They implemented genetic algorithm to solve the problem, where the fitness function considers the queue length.

The above studies have focused on determination of appropriate offsets in the design stage of signal timing plans. Certainly after implementing the timing plans in the field, there are still opportunities for fine-tuning. Gartner (2002) proposed an adaptive control strategy for synchronizing traffic signals using the virtual-fixed-cycle concept. The strategy can continuously optimize signal settings in response to demand fluctuations, which is achieved via executing a distributed dynamic programming algorithm by means of a three-layer architecture. Shoup and Bullock (1999) examined a concept of using the link travel times observed for the first vehicle in a platoon to adjust offsets. The concept could lead to an online offset refiner, if vehicle identification technologies had been deployed in arterial corridors. Abbas et al. (2001) developed an online real-time offset transitioning algorithm that continually adjusts the offsets with the objective of providing smooth progression of a platoon through an intersection. More specifically, the objective was achieved by moving the green window so that more of the current occupancy actuation histogram is included in the new window. A greedy search approach was used to determine the optimal shift of the green window. In the ACS-Lite system (2003) a run-

time refiner can modify the cycle, splits and offsets in an incremental way, based on observation of traffic conditions. Gettman et al. (2007) elaborated the data-driven algorithm in the ACS-Lite for tuning offsets. The algorithm uses upstream detectors to construct cyclic profiles of traffic arrivals and then adjusts the offsets to maximize the number of vehicles arriving during the green phase. Yin et al. (2007) proposed an offline refiner to fine-tune signal offsets, making use of a large amount of archived signal status data from real-time signal operations. Based on a more realistic estimation of the distributions of the starts/ends of green of the sync phases from the data, the refiner adjusts the offsets to minimize the red-meeting probability of the leading vehicles as well as maximizing the average bandwidth.

The systems developed so far have deficiency in considering the over-saturation phenomena such as the blockage due to an upstream queue spillback or the left-turning vehicles stored in a left-turn bay. The difficulties come from the limitations of the utilized traffic flow models, which are not able to adequately capture these over-saturation characteristics.

2.2.3 Grid Network

Intersections to be considered are often located in grid networks with either crossing arterials or a series of intersecting streets with comparable function and traffic volumes. In these situations, the entire network is often timed together. Those grid networks with short block spacing, particularly in downtown environments, are frequently timed using fixed settings and no detection (FHWA, 2008).

Many optimization approaches described in the previous section for arterial signal optimizations can be further extended to deal with the grid network problems. For examples, the multiband/multiweight approach that Gartner and Stamatiadis (2002) developed was further expanded for the network-wide coordination and a decomposition method was proposed to solve the mixed integer problem.

Recent advancements in technology allow a direct linkage between simulation models and the actual signal controllers or their software emulations, known as hardware-in-the- loop (HITL) or software-in-the-loop (SITL) (Martinez et al., 2003). These in-the-loop simulations allow actual controllers and/or their algorithms to replace the approximation used in a simulation model to more accurately reflect how a controller operates. In these systems, a simulation model first generates traffic flows and sends vehicles to the controllers via a detection design implemented in the simulation. The controller receives the calls as if it were operating in the field and then uses its own internal algorithms to adjust signal timings based on the calls received and the current signal timing. Finally, the signal displays are passed back to the simulation model, to which traffic responds.

Many commercial signal optimization packages are capable of optimizing signal plans for grid networks. Each system uses one particular algorithm or procedure to evaluate a variety of combinations of cycle length, splits, and offsets with respect to one or more performance indices and then attempts to find an optimal combination that minimize or maximize those performance indices. These systems can provide a quick response to the temporary changes in traffic pattern. SCOOT and SCATS are two examples, described as follows:

SCOOT (e.g., Hunt et al., 1981) is the abbreviation of Split Cycle Offset Optimization Technique, which is an adaptive signal optimization system. The offset selection is similar to TRANSYT, but the split at each intersection is optimized to minimize the maximum of degree of saturation at each intersection. SCOOT is computationally efficient and thus it can be implemented online, seeking small and frequent modification in the signal settings.

SCATS (Lowrie, 1982) is the abbreviation of Sydney Coordinated Adaptive Traffic System, which has been widely implemented worldwide. The system provides two levels of

optimization: strategic and tactical. Strategic control continuously optimizes the signal settings for the whole network, while the tactical control provides rapid response to the traffic conditions at the intersection level.

One issue with the adaptive signal control systems is that when demand increases, they always seek longer cycle length, which may not be preferred when maximizing throughput or minimizing queue length. Practitioners generally do not think adaptive control to be effective in congested conditions (FHWA, 2009).

2.3 Summary

This chapter has briefly reviewed the signal timing optimization literature for signal control at isolated intersections, along arterials, and in grid networks. For each type of network, previous studies are discussed according to the controller types, in the sequence of fixed-time, actuated and adaptive control.

Timing an isolated signal only needs to consider the traffic conditions at the particular intersection. Much research has been carried out for both uncongested and congested conditions, among which some have explored the methodology for timing signals under demand uncertainties. To time signals along arterials or in grid networks, another layer of complexity is to coordinate among the signals. Therefore the flow propagation between signals needs to be taken into account, which, to a great extent, influences the performance of the final signal plans. So far, few studies have been reported to directly solve the problem, let alone to consider uncertain traffic conditions.

Although adaptive systems can continuously optimize or adjust the signal setting s and provide up-to-date signal plans to deal with fluctuating traffic conditions, they are still too expensive to be widely implemented. Therefore a more proactive and cost efficient approach

needs to be developed for fixed-time and actuated signals to explicitly deal with the uncertainties associated with timing optimizations at the arterial or network level.

CHAPTER 3
A STOCHASTIC PROGRAMMING APPROACH FOR ROBUST SIGNAL TIMING OPTIMIZATION

3.1 Uncertainties in Traffic Modeling

Uncertainties arise from many aspects in planning, operating and managing a transportation system. Sources of uncertainties mainly include the randomness in demand and supply, and also travelers' stochastic and irrational behaviors (Lou et al., 2009). Addressing uncertainty in modeling transportation systems has received more and more attention recently.

3.1.1 Supply Uncertainty

Physically, a transportation system is composed of a variety of transportation facilities, whose main characteristics are represented as critical parameters in traffic modeling, such as capacity, the saturation flow rate and free-flow travel time. These parameters are stochastic in nature. Take capacity as an example. In traditional procedures for assessing of the level of service for highway facilities, capacity is treated as a constant value. For example, the design capacities of freeway facilities are derived using some given guidelines based on speed-flow diagrams like in Highway Capacity Manual (Geistefeldt, 2008). Ponzlet (1996) demonstrated that highway capacities vary according to external conditions and have to be regarded as random variables. Among others, the differences in the individual driver behaviors, vehicle characteristics, changing road and whether conditions, all contribute to variations in capacity values.

A few models are available to estimate a single value of the capacity or its distribution, like bimodal distribution method, selected maxima method, expected extreme value method, fundamental diagram method and such (Minderhoud et al., 1997; Brilon et al., 2005). To

estimate the capacity distribution, both nonparametric and parametric methods can be used, among which Weibull distribution is validated by many real data.

3.1.2 Demand Uncertainty

Demand uncertainty is one of the major uncertainties in traffic modeling. The travel demands and traffic arrivals to intersections can vary significantly even for the same time of day and day of week. As an example, Figure 3-1 displays the hourly arrivals at two crossing streets, 34th Street and University Avenue, in Gainesville, Florida, during an AM peak on weekdays over a period of four months. The flows present significant day to day variations.

The uncertainty of the demand data used in modeling can arise from two aspects: 1) the stochastic nature of the demand, 2) the detection or prediction errors. Like most of the studies in the literature, this dissertation mainly considers the uncertainty from the demand side. The uncertainty from the supply side, like stochastic capacity, can be incorporated in a similar way into the modeling framework.

3.2 Optimization under Uncertainty

Problems of optimization under uncertainty appear in many disciplines like finance and manufacturing, other than transportation engineering. Many techniques have been developed to address uncertainty, and this section introduces three widely-used ones: simulation optimization, stochastic programming, and robust optimization.

3.2.1 Simulation Optimization

Simulation optimization assumes almost nothing about the structure of the model to be optimized: uncertain parameters may depend on decision variables and the objective and constraints may be non-linear, non-convex or non-smooth functions of the variables. At each iteration, the simulation-based optimizer proposes values for the decision variables, followed by

simulation runs to evaluate the objective and constraints. Since the derivative information is hard to obtain, many trial values are needed. For each set of trial values, a thousand or more simulation runs may be conducted. In some applications, the optimization model is truly a "black box" with no discoverable structure for the problem. The simulation optimization seems the best choice for those applications (Rockafellar, 2001).

Simulation optimization is often slow, even on relatively small problems involving tens to hundreds of decision variables and uncertain parameters. Often such applications are left to run overnight or over days.

3.2.2 Stochastic Programming

Stochastic programming models take advantage of the fact that the uncertain elements in a problem can be modeled as random variables to which the theory of probability can be applied. Thus, the probability distributions of the random elements must be available (Rockafellar, 2001). Moreover, the probability distributions remain unaffected by any of the decisions taken during the optimization process.

Stochastic programming typically seeks to find some policy or plan that is feasible for all (or almost all) the possible data instances and minimizes expected values of the optimal cost function based on the known probabilistic constraints, which are often approximated via a finite scenario approach. The technique has been successfully applied in many areas, however it still remains challenging to implement for several reasons: first, the probability distributions of uncertain parameters are hard to obtain (or guess, or estimate); second, the scenario approach becomes quickly computationally intractable as the problem size grows (Sahinidis, 2004; Rockafellar, 1991).

3.2.3 Robust Optimization

While applications of stochastic programming have been reported over many years in the literature, robust optimization has gain more recent attention.

Robust optimization describes uncertainties via bounded sets, as opposed to the probability distributions for stochastic programming. With few clue to obtain the whole probability distributions, the uncertain parameters are assumed to belong to some bounded sets. The robust counterpart usually minimizes the worst-case value of the cost function to obtain a solution to the problem. Just as the stochastic programming approach, robust optimization often leads to computationally demanding problems. However, the use of sets rather than probability distributions makes it easier to use the arsenal of convex analysis, to derive approximation to the problem, as well as bounds on the quality of the approximation. (Ben-Tal et al., 2002; Bertsimas, 2003)

3.3 Scenario-Based Stochastic Programming Approach for Signal Optimization

As the advancement of portable-sensor and telecommunications technologies make high-resolution traffic data more readily available, the probability distributions of the uncertain parameters of interest become easier to estimate. Thus the stochastic programming approach can be applied to address the uncertainties associated with signal timing optimization.

Current traffic signal control strategies assume the traffic pattern will not change, which is reasonable if the solution being sought is just for next few seconds or minutes. Though clearly it is not reasonable to assume tomorrow will be just as same as today, considering the huge variance shown in Figure 3-1.

A consequent issue that traffic engineers may be confronted with is to determine what flows to use to optimize signal timings. Use of the average flows (i.e., q^0 in Figure 3-1) may not

be a sensible choice. Heydecker (1987) pointed out that if the degree of variability of traffic flows is significant, optimizing signal timing with respect to the average flows may incur considerable additional delay, compared with the timing obtained by taking this variability into account. If the degree of variability is small, use of the average flows in conventional timing methods will only lead to small losses in average performance (efficiency). However, as observed in a preliminary investigation (Yin, 2008b), it may still cause considerable losses in the performance against the worst-case scenarios or the stability of performance (robustness), thereby causing motorists' travel time to be highly variable. On the other hand, if the highest observed flows are used instead, the resulting timing plans may be over-protective and unjustifiably conservative. The average performance is very likely to be inferior. Smith et al. (2002) suggested using 90^{th} percentile volumes as the representative volumes to generate optimal timing plans and further noted that if time permits, other percentile volumes should be used to compare the results. However, it is well known in the statistical literature that extreme value estimates can be easily biased and highly unreliable when not computed properly (Reiss and Thomas, 2007).

This dissertation is to develop a methodology to fully utilize the field data, e.g., the collected traffic flow data here, to design a robust optimal signal timing plan. The performance of such timing plan is optimal against worst-case traffic conditions and also stable under any realization of uncertain traffic flows in arterials. As a preliminary investigation, Yin (2008b) developed two robust timing approaches for isolated fixed-time signalized intersections. The first approach assumes specific probabilistic distributions of traffic flows and then formulates a stochastic programming model to minimize the mean of the delays exceeding the α -percentile (e.g., 90^{th} percentile) of the entire delay distribution. In contrast, the second approach assumes

uncertain traffic flows to be unknown but bounded by a likelihood region, and then optimizes signal timing against the worst-case scenario realized within the region. It has been demonstrated that, when compared with traditional timings, robust timings may reduce the worst-case delay per vehicle by 4.9% and 11.3% respectively as well as the standard deviation of delay per vehicle by 12.0% and 16.3% respectively, without adversely affecting the average performance at a real-world intersection.

This dissertation extends the first approach, i.e., stochastic programming approach, to a general setting, optimizing the timings of fixed-time or actuated signals along arterials. It fully recognizes the uncertainty of traffic flows and assumes that they follow certain probability distributions. To represent the uncertainty of traffic flows, a set of scenarios $\Omega = \{1,2,3,\Lambda\ ,K\}$ is introduced. For each scenario $k \in \Omega$, the probability of occurrence is p_k. With these random generated scenarios, it is feasible to formulate a stochastic program to maximize the average performance of the robust signal plan across all scenarios. However, practically travelers and system managers may be more concerned with the adverse system performance, and are less likely to complain if system performs better than expected. To address such a risk-averse attitude and avoid being too conservative, we attempt to determine a robust timing plan that performs better against high-consequence or worst-case scenarios. More specifically, we minimize the expected loss (to be defined. Loss may represent different things in different settings, e.g., delay) incurred by those high-consequence scenarios whose collective probability of occurrence is $1-\alpha$, where α is a specified confidence level (say, 80%). In financial engineering, the performance measure is known as conditional value-at-risk (CVaR), or mean excess loss (Rockafellar and Uryasev, 2000). See Figure 3-2A for an illustration of the concept. The probability density function of a continuous regret and the probability mass function for discrete

case are shown in the figure. The right tail area has an area size 1-α, which contains relatively higher losses. And the CVaR is simply the mean of the losses in this area. By minimizing the CVaR, it can be claimed that the losses incurred by the high-consequence scenarios are minimized.

The following is an example to further illustrate the CVaR concept: 15 scenarios with equal probability of occurrence (1/15) are generated to represent the demand uncertainty. The corresponding delays resulted from the same signal settings are listed in Table 3-1. Choose confidence level α = 80%, then $15*(1-\alpha)=3$ scenarios with the largest delays are regarded as the high-consequence scenarios, which are scenarios 5, 13 and 1, with corresponding delays 0.89, 0.92 and 0.95. Then the CVaR value can be easily calculated as

$$\frac{1}{1-0.8}\cdot(0.89*1/15+0.92*1/15+0.95*1/15)=0.92.$$

Generally, for each scenario *k* and one particular feasible signal plan, the loss can be computed according to the problem settings, which we denote as L_k. Consider all the scenarios and order the loss as $L_1 < L_2 < ... < L_k$, let k_α be the unique index such that:

$$\sum_{k=1}^{k_\alpha} p_k \geq \alpha > \sum_{k=1}^{k_\alpha - 1} p_k$$

In words, L_{k_α} is the maximum loss that is exceeded only with probability $1-\alpha$, called as α-value-at-risk. Consequently, the expected loss exceeding the α-value-at-risk, i.e., conditional value-at-risk is:

$$\phi_\alpha = \frac{1}{1-\alpha}[(\sum_{k=1}^{k_\alpha} p_k - \alpha)L_{k_\alpha} + \sum_{k=k_\alpha+1}^{K} p_k L_k] \quad (3\text{-}1)$$

The second component in the bracket is simply to compute the mean value, and the first is to split the probability 'atom' at the delay point L_{k_α} to make the collective probability of scenarios considered in the bracket exactly equal to $1-\alpha$. See Figure 3-2B for an illustration of the concept. It can be seen that the probability mass function has a jump at the point L_{k_α} due to the associated probability of p_{k_α}, which makes $\sum_{k=1}^{k_\alpha} p_k > \alpha$. To make the collective probability of scenarios exactly equal to $1-\alpha$, we need to split the probability of delay L_{k_α}. Note that if p_{k_α} makes $\sum_{k=1}^{k_\alpha} p_k = \alpha$, then "split" is not needed, and Equation (3-1) reduces to

$$\phi_\alpha = \frac{1}{1-\alpha}\left(\sum_{k=k_\alpha+1}^{K} p_k L_k\right).$$

For each feasible signal plan, Equation (3-1) can be used to compute the resulting conditional value-at-risk and our intention is to find a signal plan that leads to the minimum conditional value-at-risk. Rockafellar and Uryasev (2002) showed that minimizing Equation (3-1) is equivalent to minimizing the following equation:

$$\min_{t,\xi} Z_\alpha = \xi + \frac{1}{1-\alpha}\sum_{k=1}^{K} p_k \cdot \max\left(L_k(t)-\xi, 0\right)$$

where ζ is a free decision variable. Subjective to a set of specific constraints, the optimal value of the objective function is the minimum conditional value-at-risk and the optimal solution $\left(t^*, \zeta^*\right)$ represents the robust signal timing plan and α-value-at-risk respectively.

The proposed stochastic programming framework is implemented in the next three chapters. Noted that, owing to a large number of notations required in each application, the notations defined in each chapter only apply to that particular chapter.

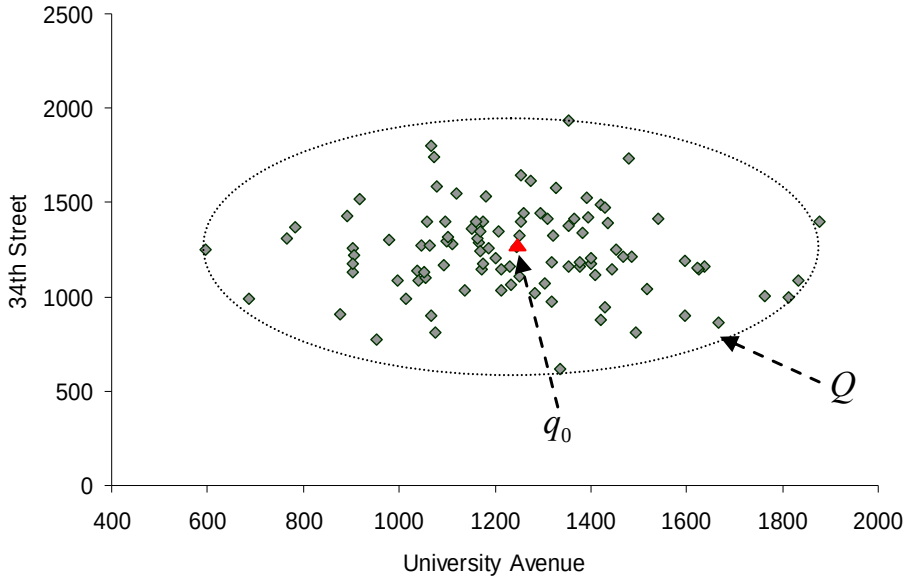
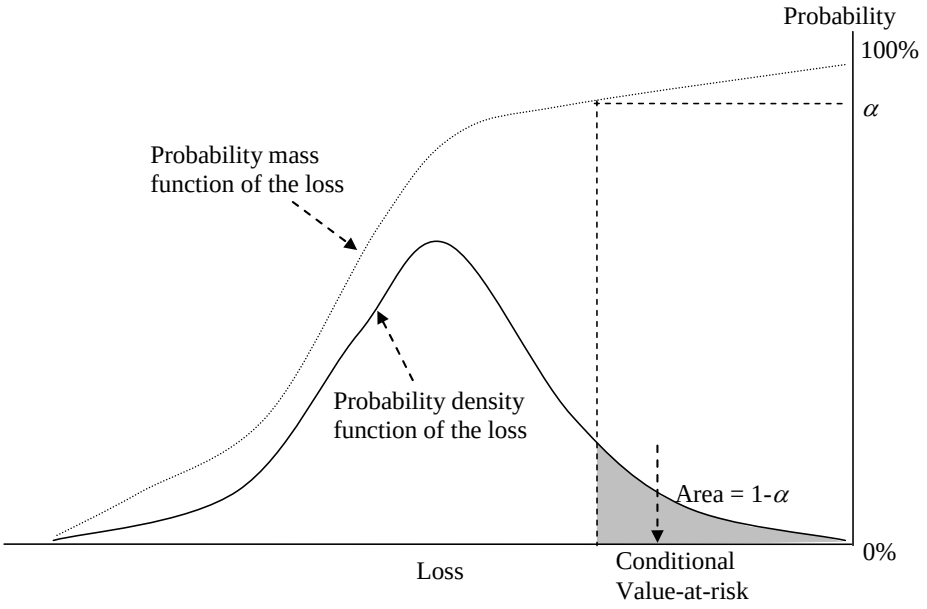
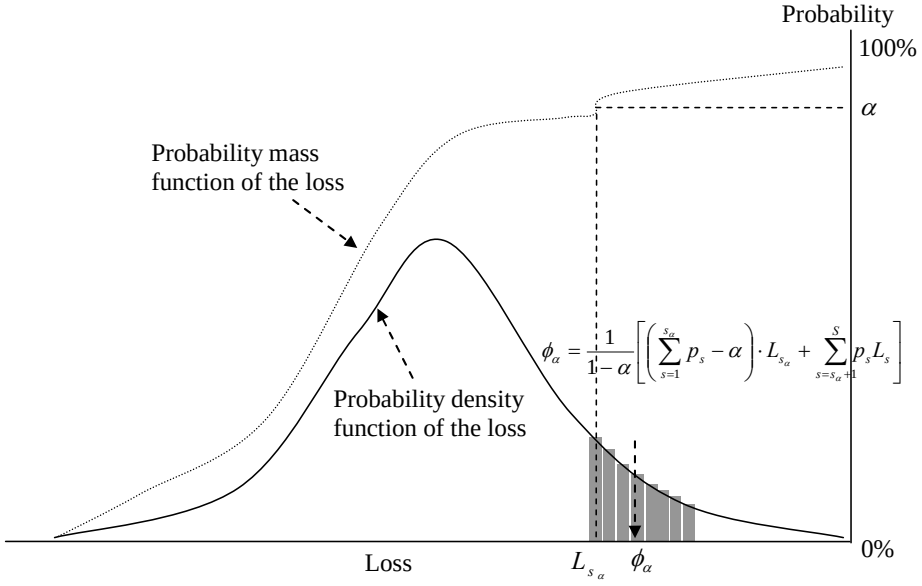


Figure 3-1. AM-peak hourly flow rates at one intersection in Gainesville, FL



A



B

Figure 3-2. Illustration of the concept of conditional value-at-risk. A) A continuous loss function. B) A discrete loss function.

Table 3-1. Delay associated with 15 scenarios

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.95	0.23	0.61	0.49	0.89	0.76	0.45	0.01	0.82	0.44	0.61	0.79	0.92	0.73	0.17

CHAPTER 4
SIGNAL TIMING OPTIMIZATION UNDER DAY-TO-DAY DEMAND VARIATIONS

Since the seminal work of Webster (1958), significant efforts have been devoted to improving signal timing for saturated isolated intersections, coordinated arterials and grid networks etc. However, only a few studies have been conducted in the literature to directly address signal timing under flow fluctuations for fixed-time control systems.

This chapter applies the stochastic programming approach to optimize the timings of signals along arterials under day-to-day demand variations or uncertain traffic future growth. Based on a cell-transmission representation of traffic dynamics, a stochastic programming model is formulated to determine cycle length, green splits, phase sequences and offsets to minimize the expected delay incurred by high-consequence scenarios of traffic demand. The stochastic programming model is simple in structure but contains a large number of binary variables. Existing algorithms, such as branch and bound, are not able to solve it efficiently. Consequently, a simulation-based genetic algorithm is developed to solve the model.

4.1 Cell-Transmission Model

4.1.1 Model Introduction

Modeling traffic dynamics is particularly important for signal timing optimization because realistic evaluation of each feasible timing plan cannot be performed without a realistic traffic flow model. At the same time, the evaluation should be efficient such that it can be incorporated into an optimization procedure. For these reasons, we select the macroscopic cell-transmission model (CTM) proposed by Daganzo (1994, 1995) in order to fully capture traffic dynamics, such as shockwaves, and queue formation and dissipation.

CTM is a finite difference solution scheme for the hydrodynamic theory of traffic flow or the Lighthill-Whitham-Richards (LWR) models. Mathematically the theory can be stated as the following equations:

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \tag{4-1}$$

$$q = Q(k, x, t) \tag{4-2}$$

Where the q and k are two macroscopic variables: flow and density. Equation (4-1) is the flow conservation equation and Equation (4-2) defines the traffic flow (q), at location x and time t, as a function of the density (k).

For a homogeneous roadway, Daganzo (1994, 1995) suggested using the time-invariant flow-density relationship:

$$q = \min\{Vk, Q, W(k_{jam} - k)\}$$

where V = the free flow speed;

Q = the inflow capacity;

k_{jam} = the jam density;

W = the backward wave speed.

Figure 4-1 shows the flow-density relationship in a piecewise linear diagram.

By dividing the whole network into homogeneous cells with the cell length equal to the duration of time step multiplied by the free flow speed, the results of the LWR model can be approximated by a set of recursive equations:

$$n_i(t+1) = n_i(t) + y_{i-1}(t) - y_i(t) \tag{4-3}$$

$$y_i(t) = \min\{n_i(t), Q_i(t), \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]\} \tag{4-4}$$

where $n_i(t)$ = the number of vehicles in cell i during time step t;

$y_i(t)$ = the number of vehicles that leave cell i during time step t;

$N_{i,\max}$ = the maximum number of vehicles that can be accommodated by cell i;

$Q_i(t)$ = the minimum of the capacity flows of cell i and $i+1$;

$\omega = W / V$.

Equation (4-3) ensures the flow conservation that the number of vehicles in cell i during time step $t+1$ equals to the number of vehicles in cell i during time step t plus the inflow and

minus the outflow. Equation (4-4) determines the outflow for each cell during each time step, which is a piecewise linear function.

4.1.2 Encapsulating CTM in Signal Timing Optimization

Lin and Wang (2004), Lo (1999) and Lo et al. (2001) have successfully incorporated CTM in their signal timing optimization formulations.

Lin and Wang (2004) formulated a 0-1 mixed integer linear program, considering the number of stops and fixed or dynamic cycle length. In the model, cells in the network are categorized into four groups: ordinary, intersection, origin and destination cells. The objective is to minimize a weighted sum of total delay and total number of stops. In their model, Equation (4-4) is replaced by three linear inequalities, which do not accurately replicate flow propagation and may suffer the so-called “vehicle holding problem”. To address this issue, one additional penalty term is added to the objective function. The authors demonstrated the model capable of capturing traffic dynamic using an emergency vehicle problem. However, the model is developed only for one-way streets and neither merge nor diverge of traffic is considered.

Lo (1999, 2001) and his colleagues (2001) developed dynamic signal control formulations based on CTM. By introducing binary variables, Equation (4-4) is equivalently converted into a linear system. The models proposed are able to generate dynamic or fixed timing plan and optimize cycle length, phase splits and offsets explicitly. Unfortunately, the models are again proposed for one-way streets.

This section expands Lo’s models to a more general and realistic setting, including modeling two-way traffic, phase sequence optimization and applying new technique to equivalently transforming CTM for a general signal-controlled network to be a linear system of equalities and inequalities with integer variables.

4.2 Enhanced Deterministic Signal Optimization Model

Assuming deterministic constant or time-variant demands within the optimization horizon, this section presents a CTM-based deterministic signal timing optimization model. The model extends Lo's models in the following aspects:

- Modeling two-way traffic: this extension not only increases the size of the problem but also introduces another layer of complexity in representing signalized intersections and signal settings. For example, as the number of traffic movements increases, the number of phase combinations and sequences increase significantly;
- Optimization of phase sequence: the left-turn leading or lagging control is modeled explicitly;
- New formulation: we transform the CTM of a general signal-controlled arterial to be an equivalent linear system of equalities and inequalities with integer variables using a technique recently proposed by Pavlis and Recker (2009) and formulate a mixed-integer linear program to optimize cycle length, green splits, offsets and phase sequences.

It is assumed that every intersection along the arterial is signalized, and all the cells comprising the network can be categorized into six groups: ordinary, origin, destination, non-signalized diverge, signalized diverge and signalized merge cells, as shown in Figure 4-2 from A to F. Each group has a different configuration to be discussed below.

4.2.1 Objective Function

In the deterministic setting, we aim to optimize signal timing to minimize the total system delay of an urban arterial. The objective is to minimize the total area (as in Figure 4-3) between the cumulative arrival curves of the origin cells and the cumulative departure curves of the destination cells, expressed as the following linear function:

$$L = \min\left(\sum_{i\in O}\sum_{t=1}^{T}\sum_{j=1}^{t} d_i(j)\right) - \sum_{i\in D}\sum_{t=1}^{T}\sum_{j=1}^{t} y_i(j)\right)$$

where, O is the set of origin cells and D is the set of destination cells; T is the duration of the optimization horizon; $d_i(j)$ is the demand at origin cell i during time step j. It is straightforward to observe that if the demands at origin cells are given, the objective function is

equivalent to maximizing the second component, i.e., the area under the cumulative departure curves.

4.2.2 Constraints

4.2.2.1 Constraints for Ordinary Cells

The ordinary cells are those with only one inflow and one outflow as cell i in Figure 4-2A. According to the cell transmission model, the flow constraints are as follows:

$$n_i(t+1) = n_i(t) + y_{i-1}(t) - y_i(t)$$

$$y_i(t) = \min\{n_i(t), Q_{i,\max}(t), Q_{i+1,\max}(t), \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]\}$$

$y_i(t)$ is determined by the min function, which is essentially a linear conditional piecewise function (CPF). Pavlis and Recker (2009) provided a scheme to transform this kind of CPF into mixed integer constraints with the least number of integer variables. By introducing two binary variables, i.e., χ_1 and χ_2, and a sufficiently large negative constant, i.e., U^-, the CPF can be equivalently translated into the following constraints:

$$(\chi_1 + \chi_2)U^- \le y_i(t) - n_i(t) \le 0$$

$$(1 + \chi_1 - \chi_2)U^- \le y_i(t) - Q_{i,\max}(t) \le 0$$

$$(1 - \chi_1 + \chi_2)U^- \le y_i(t) - Q_{i+1,\max}(t) \le 0$$

$$(2 - \chi_1 - \chi_2)U^- \le y_i(t) - \omega \cdot [N_{i+1,\max} - n_{i+1}(t)] \le 0$$

To see the equivalence, Table 4-1 enumerates all possible 0-1 combinations and evaluates the value of $y_i(t)$.

4.2.2.2 Constraints for Origin Cells

The origin cells, as shown in Figure 4-2B, have the same structure as the ordinary cells, except that the inflow is fixed as the corresponding demand input. These cells perform as valves

that control the traffic volume flowing into the network. The above constraints are slightly changed to incorporate the demand:

$$n_i(t+1) = n_i(t) + d_i(t) - y_i(t)$$

$$(\chi_3 + \chi_4)U^- \le y_i(t) - n_i(t) \le 0$$

$$(1 + \chi_3 - \chi_4)U^- \le y_i(t) - Q_{i,\max}(t) \le 0$$

$$(1 - \chi_3 + \chi_4)U^- \le y_i(t) - Q_{i+1,\max}(t) \le 0$$

$$(2 - \chi_3 - \chi_4)U^- \le y_i(t) - \omega \cdot [N_{i+1,\max} - n_{i+1}(t)] \le 0$$

4.2.2.3 Constraints for Destination Cells

The destination cells, as in Figure 4-2C, are those with outflow unlimited, implying that all the vehicles currently reside in the cells are able to flow out of the system at the next time step. The constraints are as follows:

$$n_i(t+1) = n_i(t) + y_{i-1}(t) - y_i(t)$$

$$y_i(t) = n_i(t)$$

4.2.2.4 Constraints for Non-Signalized Diverge Cells

Non-signalized diverge occurs at certain roadway segments where the geometry or capacity changes and traffic diverse to different lanes for their respective destinations. Figure 4-2D is a typical configuration for non-signalized diverge: traffic in cell i diverges to cells j_1 and j_2 according to proportion parameters β_{j_1} and β_{j_2}. The constraints can be stated as follows:

$$y_{j_1-1} = \beta_{j_1} \cdot y_i$$

$$y_{j_2-1} = \beta_{j_2} \cdot y_i$$

$$\beta_{j_1} + \beta_{j_2} = 1$$

$$(\chi_5 + \chi_6 + \chi_7)U^- \le y_i(t) - n_i(t) \le 0$$

$$(1 + \chi_5 + \chi_6 - \chi_7)U^- \le y_i(t) - Q_{i,\max}(t) \le 0$$

$$(1 + \chi_5 - \chi_6 + \chi_7)U^- \le y_i(t) - Q_{j_1,\max}(t)/\beta_{j_1} \le 0$$

$$(1-\chi_5+\chi_6+\chi_7)U^- \le y_i(t)-Q_{j_2,\max}(t)/\beta_{j_2} \le 0$$
$$(2-\chi_5-\chi_6+\chi_7)U^- \le y_i(t)-\omega\cdot[N_{j_1,\max}-n_{j_1}(t)]/\beta_{j_1}] \le 0$$
$$(2-\chi_5+\chi_6-\chi_7)U^- \le y_i(t)-\omega\cdot[N_{j_2,\max}-n_{j_2}(t)]/\beta_{j_2}] \le 0$$
$$\chi_5+\chi_6+\chi_7 \le 2$$
$$\chi_6+\chi_7 \le 1$$

4.2.2.5 Constraints for Signalized Diverge Cells

Signalized diverge is the diverge that happens within a signalized intersection, when traffic from one direction enters the intersection during a corresponding green phase and leaves the intersection while diverging into two or more bounds of traffic. Figure 4-2E sketches a configuration of the signalized diverge, where the sign *S* indicates a traffic signal. The constraints are as follows:

$$y_{j_1-1}=\beta_{j_1}\cdot y_i$$
$$y_{j_2-1}=\beta_{j_2}\cdot y_i$$
$$\beta_{j_1}+\beta_{j_2}=1$$

$$(\chi_8+\chi_9+\chi_{10})U^- \le y_i(t)-n_i(t) \le 0$$
$$(1+\chi_8+\chi_9-\chi_{10})U^- \le y_i(t)-Q_i(t) \le 0$$
$$(1+\chi_8-\chi_9+\chi_{10})U^- \le y_i(t)-Q_{j_1,\max}(t)/\beta_{j_1} \le 0$$
$$(1-\chi_8+\chi_9+\chi_{10})U^- \le y_i(t)-Q_{j_2,\max}(t)/\beta_{j_2} \le 0$$
$$(2-\chi_8-\chi_9+\chi_{10})U^- \le y_i(t)-\omega\cdot[N_{j_1,\max}-n_{j_1}(t)]/\beta_{j_1}] \le 0$$
$$(2-\chi_8+\chi_9-\chi_{10})U^- \le y_i(t)-\omega\cdot[N_{j_2,\max}-n_{j_2}(t)]/\beta_{j_2}] \le 0$$
$$\chi_8+\chi_9+\chi_{10} \le 2$$
$$\chi_9+\chi_{10} \le 1$$

The set of constraints is identical to those for non-signalized diverge cells except that $Q_{i,\max}(t)$ is replaced by $Q_i(t)$. The value of $Q_i(t)$ depends on the status of the signal phase associated with cell i and will be discussed later.

4.2.2.6 Constraints for Signalized Merge Cells

Figure 4-2F is an example of traffic merge under signal control. According to the signal settings, these three streams of traffic entering the intersection are associated with three individual signal phases that conflict with each other. Therefore, practically there is only one stream of traffic entering the intersection at one time step. The constraints are thus as follows:

$$y_{j-1}(t) = y_{i1}(t) + y_{i2}(t) + y_{i3}(t)$$

Approach 1:

$$(\chi_{11} + \chi_{12})U^- \le y_{i_1}(t) - n_{i_1}(t) \le 0$$
$$(1 + \chi_{11} - \chi_{12})U^- \le y_{i_1}(t) - Q_{i_1}(t) \le 0$$
$$(1 - \chi_{11} + \chi_{12})U^- \le y_{i_1}(t) - Q_{j,\max}(t) \le 0$$
$$(2 - \chi_{11} - \chi_{12})U^- \le y_{i_1}(t) - \omega \cdot [N_{j,\max} - n_j(t)] \le 0$$

Approach 2:

$$(\chi_{13} + \chi_{14})U^- \le y_{i_2}(t) - n_{i_2}(t) \le 0$$
$$(1 + \chi_{13} - \chi_{14})U^- \le y_{i_2}(t) - Q_{i_2}(t) \le 0$$
$$(1 - \chi_{13} + \chi_{14})U^- \le y_{i_2}(t) - Q_{j,\max}(t) \le 0$$
$$(2 - \chi_{13} - \chi_{14})U^- \le y_{i_2}(t) - \omega \cdot [N_{j,\max} - n_j(t)] \le 0$$

Approach 3:

$$(\chi_{15} + \chi_{16})U^- \le y_{i_3}(t) - n_{i_3}(t) \le 0$$
$$(1 + \chi_{15} - \chi_{16})U^- \le y_{i_3}(t) - Q_{i_3}(t) \le 0$$
$$(1 - \chi_{15} + \chi_{16})U^- \le y_{i_3}(t) - Q_{j,\max}(t) \le 0$$
$$(2 - \chi_{15} - \chi_{16})U^- \le y_{i_3}(t) - \omega \cdot [N_{j,\max} - n_j(t)] \le 0$$

where $Q_i(t)$ is to be discussed next.

4.2.2.7 Constraints for Connection between Signal and Flow

At signalized intersections, the capacity flow of a cell depends on the status of the corresponding signal phase, because only when this phase turns green can the traffic propagates forward or makes a turn. The capacity flow satisfies the following statement:

$$\text{If } b(p) < t \le e(p), \text{ then } Q_i(t) = s \text{ ; otherwise, } Q_i(t) = 0,$$

where s is saturation flow rate;

$b(p)$ is the beginning of green phase p;

$e(p)$ is the end of green phase p.

The above if-then relationship can be translated into a system of equalities and inequalities by introducing two binary variables $z_1(p,t)$ and $z_2(p,t)$. The system is stated as follows:

$$-U \cdot z_1(p,t) + \varepsilon \le t - e(p) \le U \cdot [1 - z_1(p,t)]$$
$$-U \cdot z_2(p,t) \le b(p) - t \le U \cdot [1 - z_2(p,t)] - \varepsilon$$
$$z_1(p,t) + z_2(p,t) - z(p,t) = 1$$
$$Q_i(t) = (z_1(p,t) + z_2(p,t) - 1) \cdot s$$
$$\sum_p (z_1(p,t) + z_2(p,t)) \le 2$$

where U is a sufficiently large positive number and ε is an arbitrary small number. The last constraint ensures that there are at most two phases that can be green at the same time.

4.2.2.8 Constraints for Signal Phase Sequence

The model intends to explicitly optimize phase sequences under the NEMA phasing structure. Two types of intersections are considered as shown in Figure 4-4: A Four-way intersection and B T-intersection.

Four-Way Intersection. Figure 4-5A illustrates the standard NEMA phasing for a four-way intersection. With the barrier in the middle, the structure can be divided into four portions and phase sequence is determined within each portion.

A binary variable is introduced for each portion as shown in Figure 4-5B. Consider phase 1 and 2 as an example. Let g denote the green time duration; o denote the offset point of each signal; l be the cycle length; k indicate the signal identification number; c represent the cycle identification number and h be the barrier time point. The following constraints are included for determining the phase sequence for phase 1 and 2:

$$b(m,"1",c) = \lambda_1 \cdot o(m) + \lambda_1 \cdot l \cdot (c-1) + (1-\lambda_1) \cdot e(m,"2",c)$$
$$e(m,"1",c) = b(m,"1",c) + g(m,"1")$$
$$b(m,"2",c) = (1-\lambda_1) \cdot o(m) + (1-\lambda_1) \cdot l \cdot (c-1) + \lambda_1 \cdot e(m,"1",c)$$
$$e(m,"2",c) = b(m,"2",c) + g(m,"2")$$
$$g(m,"1") + g(m,"2") = h(m)$$

It can be seen that when λ_1 equals 1, phase 1 starts at the offset point of the signal and phase 2 follows phase 1. It is also true reversely. Similarly, by introducing another three binary variables λ_2, λ_3 and λ_4 respectively, constraints can be constructed to determine phase sequence for the pairs of phase 3 and 4, 5 and 6, 7 and 8. Once the value of (λ_1, λ_2, λ_3, λ_4) is determined, the left-turn leading and lagging information can be obtained explicitly. Figure 4-5B presents a particular phase sequence corresponding to $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = 0$ and $\lambda_4 = 0$.

T-Intersection. T-intersection can be modeled the same way as the four-way intersection but is much simpler. Figure 4-6 illustrates the NEMA phasing structure of a particular T-intersection where the only phase sequence needs to be determined is between phase 5 and 6. Therefore one binary variable λ_3 is introduced for the whole intersection. Correspondingly, only one set of constraints is needed for the entire structure, listed as follows:

$$b(k,"5",c) = \lambda_3 \cdot o(k) + \lambda_3 \cdot l \cdot (c-1) + (1-\lambda_3) \cdot e(k,"6",c)$$
$$e(k,"5",c) = b(k,"5",c) + g(k,"5")$$
$$b(k,"6",c) = (1-\lambda_3) \cdot o(k) + (1-\lambda_3) \cdot l \cdot (c-1) + \lambda_3 \cdot e(k,"5",c)$$
$$e(k,"6",c) = b(k,"6",c) + g(k,"6")$$
$$g(k,"5") + g(k,"6") = h(k)$$

4.2.3 Model Formulation

Given a particular network, the cell representation should be first constructed according to the geometry and signal setting. The cells are then classified into six categories and the corresponding set of constraints can be written for each cell as previously presented. The constraints comprise a linear system with integer variables. With the linear objective function to minimize total system delay, the optimization problem is a mixed-integer linear program. One

portion of the optimal solution to the program specifies the signal timing, denoted as a vector $\left(l^*, o^*, \lambda^*, g^*\right)^T$, where l^*, o^*, λ^* and g^* are vectors of optimal cycle length, offsets, phase sequences and green splits.

4.3 Stochastic Signal Optimization Model

In the above deterministic case, traffic demand is assumed to be fixed within the optimization horizon. However in reality, demand may vary significantly, like in Figure 3-1. We assume that the demand at each origin cell follows a certain stochastic distribution. To capture the joint stochastic distribution of traffic demands, also a set of scenarios $\Omega = \{1,2,3,\Lambda\ ,K\}$ is introduced. A typical scenario consists of demand realizations at all origin cells. More specifically, a scenario is a vector $d^k = (d_1^k, d_2^k, \Lambda\ d_r^k \cdots)^T, \forall r \in O$.

For each demand scenario *k* and one particular feasible signal plan (l, o, λ, g), the total system delay can be computed, as described in the previous section 3.2. We denote the resulting delay as $L_k(l, o, \lambda, g)$. By applying the scenario-based approach, the robust signal coordination plan can be obtained by minimizing the following equation:

$$\min_{l,o,\lambda,g,\xi} Z_\alpha = \xi + \frac{1}{1-\alpha}\sum_{k=1}^{K} \pi_k \cdot \max\left(L_k(l, o, \lambda, g) - \xi, 0\right)$$

Each demand scenario requires a set of the constraints as discussed in the previous section, so the final stochastic optimization problem will include multiple sets of such constraints depending on the number of scenarios generated.

4.4 Numerical Examples

4.4.1 Simulation-Based Genetic Algorithm

The stochastic programming model formulated above is simple in structure but contains a large number of binary variables. Therefore, existing algorithms, such as branch and bound, are not able to solve it efficiently, particularly when the optimization horizon is long and the network

size is large. We thus develop a simulation-based binary genetic algorithm (GA) to solve the model. Here the "simulation-based" means that the fitness function in the GA is evaluated through macroscopic simulation using CTM.

GAs have been widely used in different fields such as engineering, economics and physics to solve problems that are not analytically solvable or cannot be solved by traditional search methods. In the transportation literature, researchers have developed GA-based solution algorithms to solve problems including equilibrium network design, dynamic traffic assignment and second-best congestion pricing and traffic control problems. In Lo et al. (2001), GA was used to solve the signal optimization problem.

GA is a global search technique. It starts from an initial group of randomly generated feasible solutions, and then employs operations like crossover and mutation to generate the new solution pools. The iteration continues until some criterion is satisfied, e.g., the maximum number of generation. The simulation-based GA proposed in this report follows the general framework of GA, and Figure 4-7 presents the flow chart of the algorithm. There are two loops: the outer loop for counting the number of generations while the inner is to track the number of individuals within each generation. Other core components of the algorithm will be discussed next.

4.4.1.1 Chromosome Configuration

The chromosome is defined according to the decision variables, which include the cycle length, offsets, phase sequences and phase splits. Each chromosome defines a solution, and only the feasible solutions can be selected as the individuals in each generation. Figure 4-8 presents an example of the chromosome that represents a three-signal arterial. There are in total 144 bits, of which the first six bits 1-6 represent the cycle length. A length of six binary numbers can

represent a decimal value from 0 to 63. If traffic dynamic is modeled at two seconds per time step, then the cycle length can vary from 0 to 126 seconds. The rest bits are equally divided into three portions, 46 bits for each signal.

Consider the first signal. Bits 7, 8, 9, and 10 define the phase sequences for the signal as discussed in Section 3.2.8. Generally, if a bit has a value of 0, then the corresponding odd phase is activated before the even phase. Otherwise, the even phase comes first. A four-way intersection will require determining values in all four bits, while a T-intersection only needs one bit information. The next seven bits, i.e., 11-17, represent the offset for the signal. A seven-bit binary number can represent a decimal number from 0 to 127. Because an offset is expressed as a percentage of the cycle length in this report, one additional constraint on the binary number is in place to ensure the feasibility of the offset. Bits 18-24 represent the barrier point, which is also expressed as a percentage of the cycle length. Another additional constraint is required as well to ensure that the newly generated barrier point stays in the current cycle.

The next four clusters of bits represent four green times G_1, G_2, G_3 and G_4 (see Figure 4-5B), which are the green durations of the phases that lead in the respective portions. G_1 and G_3 are in percentage of the barrier time while G_2 and G_4 are in percentage of the difference between cycle length and barrier time.

4.4.1.2 Fitness Evaluation

For each generation, every individual needs to be evaluated so as to decide its priority to breed the next generation. Here each individual represents a specific signal plan and the fitness evaluation is to determine its corresponding mean excess delay. More specifically, for each individual signal plan, we run a macroscopic simulation based on CTM with all demand

scenarios and calculate the corresponding system control delays. The mean excess delay can be computed and will be used to determine its priority for breading the next generation.

4.4.1.3 Probability Assignment

Generally, the smaller the mean excess delay is, the larger probability the corresponding signal plan will be chosen to breed the next generation. To calculate these probabilities, we have tested a variety of fitness functions and the following two generally show good performance:

$$\begin{cases} \dfrac{1}{(CVAR)^5} \\ \dfrac{1}{\ln(CVAR)} \end{cases}$$

The crossover probabilities are calculated proportionally to the fitness function value.

4.4.1.4 Crossover

Crossover is the main procedure to generate new chromosomes. To increase diversity, we use multi-point crossover, developed according to the chromosome structure, other than using one-point crossover. After the selection of two parents according to the crossover probability, several crossover points will be randomly generated but ensure that one is among the first six bits, which influences the cycle length, and one for each signal, which may change the setting for each signal. Therefore, if there are n signals, there will be $n+1$ crossover points in total.

4.4.1.5 Mutation

Each crossover operation will generate two offspring and the mutation operation is subsequently conducted. Mutation randomly changes the value of the bit value in the chromosomes to increase the diversity in the population, so that the GA will have the chance to find a better solution rather than stop at one local optimum. The mutation rate to be used is 5‰.

4.4.1.6 Individual Validation

New individual produced through crossover and mutation operations may not be appropriate, i.e., the corresponding timing plan may not be technically feasible. Therefore, additional constraints need be set to ensure the validity of each individual. Our algorithm mainly checks the followings:

- The individual newly added will not repeat any individual contained in the population to maintain the diversity of the population;
- An individual with a cycle length smaller than a certain value will not be considered to ensure the cycle length to be in a reasonable range;
- The offsets must be smaller than the cycle length;
- The barrier points must be in the corresponding cycle;
- And each signal phase maintains a certain minimum green.

4.4.2 Numerical Example I

4.4.2.1 Test Network and Demand Data

The first numerical experiment is carried out on an artificial arterial with three intersections, whose cell representation is shown in Figure 4-9. The design speed limit is 35mph, which is approximately equivalent to 50 feet/second. Since traffic dynamics is modeled in a resolution of two seconds per time step, 100 feet is the cell length for all 117 cells. We implement the stochastic signal timing model under both uncongested and congested traffic conditions. The low demand in Table 4-2 is for the uncongested situation while the high demand is for the congested cases. The turning percentages at each signal are also given in the table

4.4.2.2 Plan Generation

For the comparison purpose, two plans are generated under both uncongested and congested conditions: one is called as robust plan, which is generated by solving the stochastic signal timing model with the demand scenarios created from the uniform distributions shown in Table 4-2; the other is called nominal plan, derived by solving the deterministic signal timing model with the mean demand. Because the two fitness functions have similar convergence speed

and generate timing plans with similar performance, we only report the result from using the log-form fitness function.

According to the convergence performance of the algorithm, we set the maximal number of generations to 600 and 1000 for the uncongested and congested case respectively. Figure 4-10 presents the convergence tendency of the algorithm in both cases. The algorithm converges faster in the uncongested case, particularly in the early stage of the iterations.

Table 4-3 presents the resulting signal plans. The phase sequence is given in form of binary vector while others are decimal numbers in the unit of second. The minimum green for each phase is set as four seconds. P1 to P8 stand for the phases in NEMA phasing.

4.4.2.3 Plan Evaluation

We compare the robust and nominal signal plans using the microscopic CORSIM simulation, and the system delay is selected as the performance measure. Figure 4-11 is a snapshot of the CORSIM network. In the simulation, demand scenarios are obtained by sampling the uniform distributions provided in Table 4-2. Table 4-4 summarizes the simulation result. It can be seen that the robust timing plans reduce the mean excess delay by 28.68% in the uncongested case and 7.46% in the congested case. In both cases, it also improves the average delay across all demand scenarios by over 20%.

4.4.3 Numerical Example II

4.4.3.1 Test Network and Demand Data

The second numerical experiment is carried out on a stretch of El Camino Real in the San Francisco Bay Area of California, starting from Crystal Springs Rd to 5th Ave. Figure 4-12 is the cell representation of the arterial. The speed limit on the major street is 35 mph or 50 feet per second while 25 mph or 36 feet per second on the side streets. Because traffic dynamics is modeled second by second, the cell length for the major and side streets is 50 and 36 feet

respectively. Traffic demand data were collected from loop detectors for peak hours in a duration of 10 working days in July 2008. Table 4-5 provides a summary of the flow data and the turning proportions at the intersections.

4.4.3.2 Plan Generation

The observed flow rates are used directly as demand scenarios with equal probability of occurrence, to generate the robust plans by solving the stochastic programming model using the simulation-based GA approach. For comparison, a nominal plan is generated by solving the deterministic model with the mean demands presented in Table 4-5. Both robust and nominal plans are generated after 600 generations. Figure 4-13 shows the convergence of the GA with both 5^{th}-form and log-form fitness functions. It can be observed the 5^{th}-form fitness function converges faster than the log-form counterpart.

Table 4-6 presents the signal plans generated from both fitness functions, and their performance will be compared next. The minimum green for each phase is set as eight seconds.

4.4.3.3 Plan Evaluation

The comparison is also conducted via microscopic simulation with demand profiles randomly generated based on Table 4-5 assuming truncated normal distributions. Figure 4-14 is a snapshot of the CORSIM network for the corridor. Table 4-7 presents the CORSIM simulation result. The traffic condition is very congested through the whole simulation period. It can be seen that the robust plans outperform the corresponding nominal plan, with the mean delay reduced by 23.69% and 17.82%, and the mean excess delay reduced by 22.80% and 17.34%. It demonstrates that the robust plans perform much better against high-consequence scenarios. As a side effect, the average performance is also improved. Although the 5th-form fitness function

leads to a faster convergence, it does not improve the performance as much as the log-form fitness function does.

4.5 Summary

This chapter has presented a stochastic model to optimize traffic signals on arterials considering day-to-day demand variations or uncertain further demand growth. The model optimizes the cycle length, green splits, offset points and phase sequences in an integrated manner. The resulting robust timing plans have been demonstrated to perform better against high-consequence scenarios without losing optimality in the average sense.

Considering a large number of binary variables in the formulation, we have developed a simulation-based GA to solve the problem. It should be mentioned that the setting of the GA-based algorithm, such as the fitness function, may influence the quality of the final plan and the convergence speed. Numerical experiments are needed to fine-tune the setting. We also note that the simulation-based model is broadly applicable, particularly when the objective function is difficult or time-consuming to evaluate.

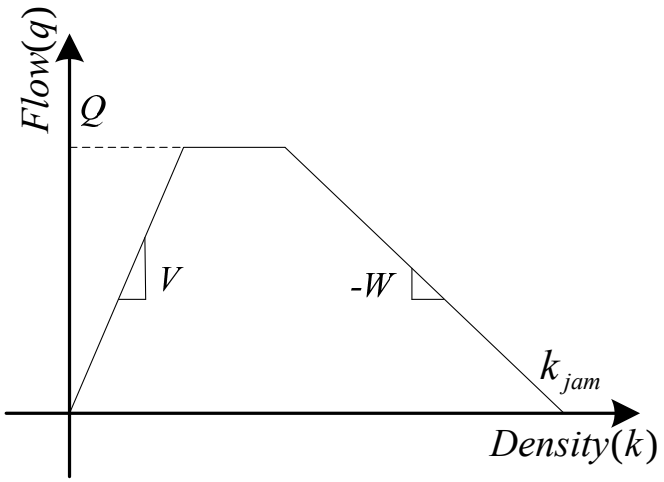


Figure 4-1. Piecewise linear $q-k$ relationship

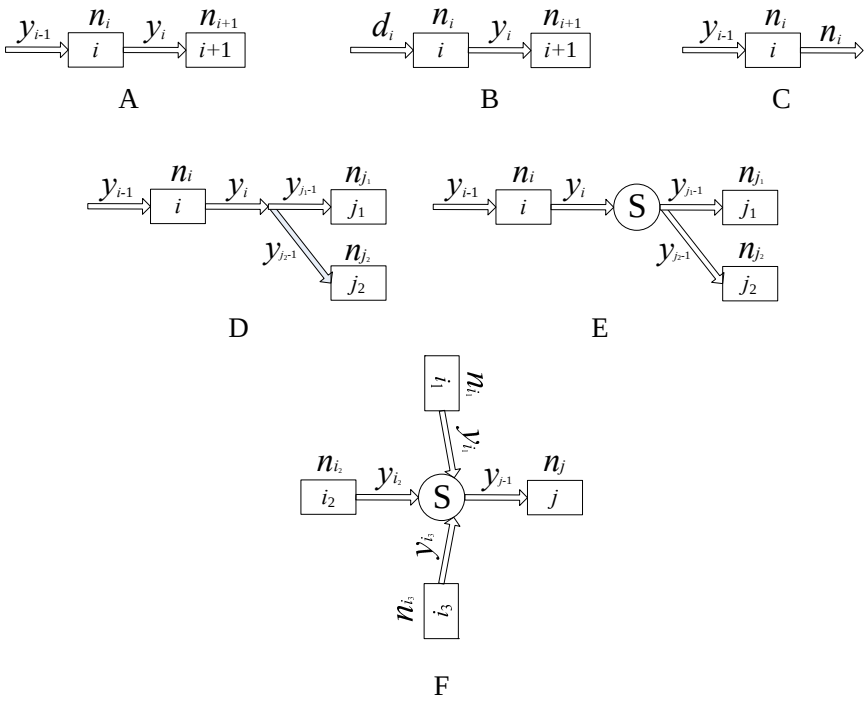


Figure 4-2. Cell configurations

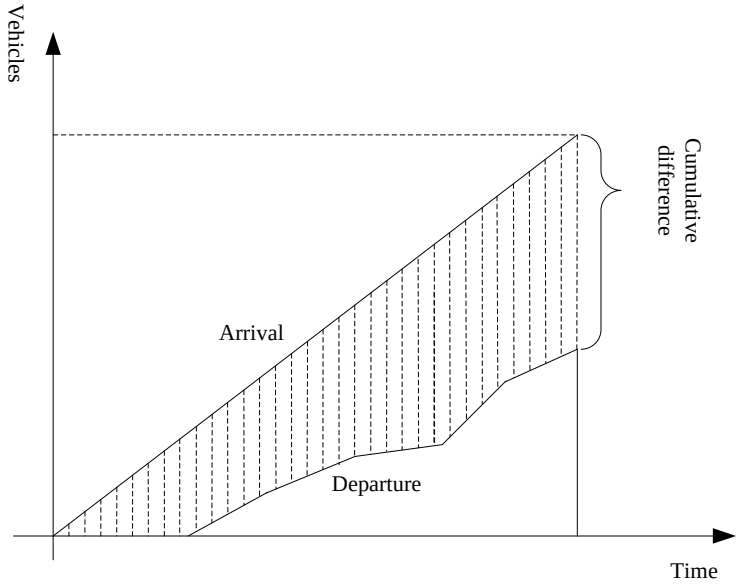


Figure 4-3. Interpretation of the objective function

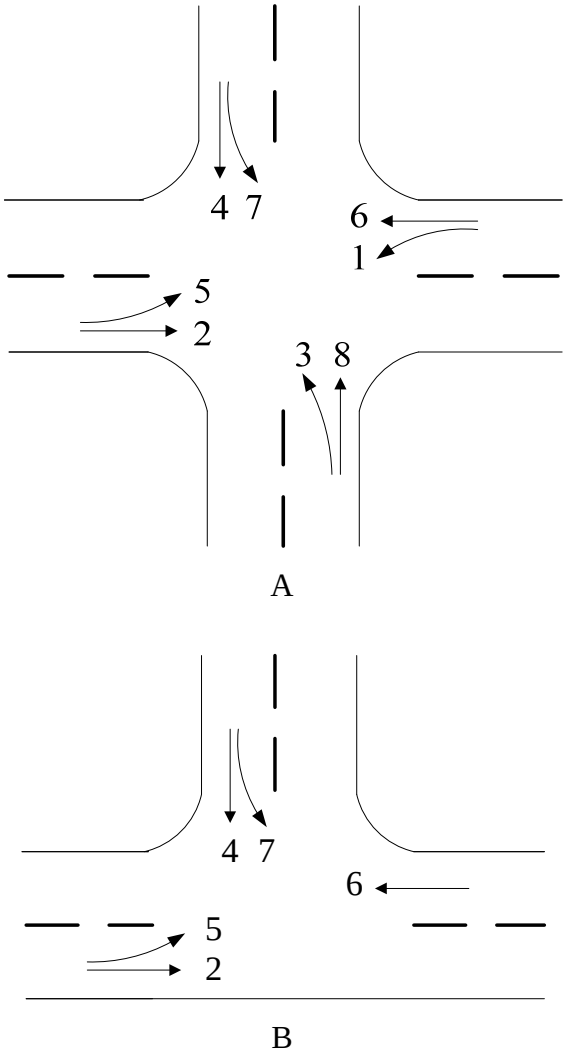
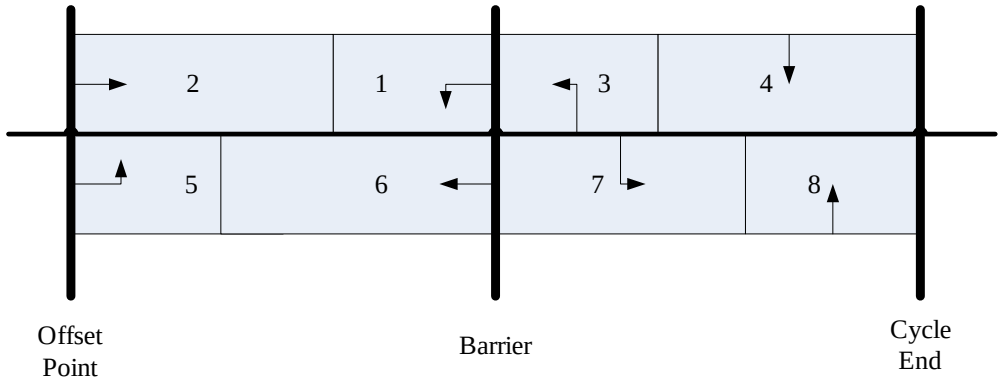
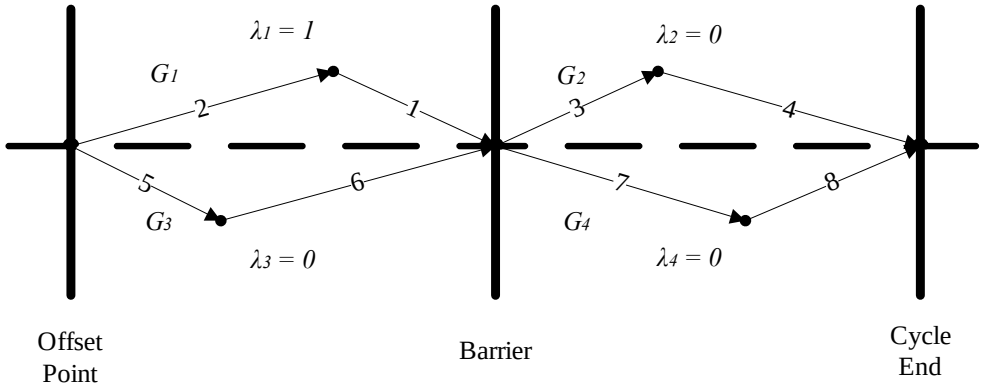


Figure 4-4. Traffic intersection configurations. A) Four-way intersection. B) T-intersection.



A



B

Figure 4-5. NEMA phasing structure for a four-way intersection. A) NEMA phasing. B) Transformation of NEMA phasing.

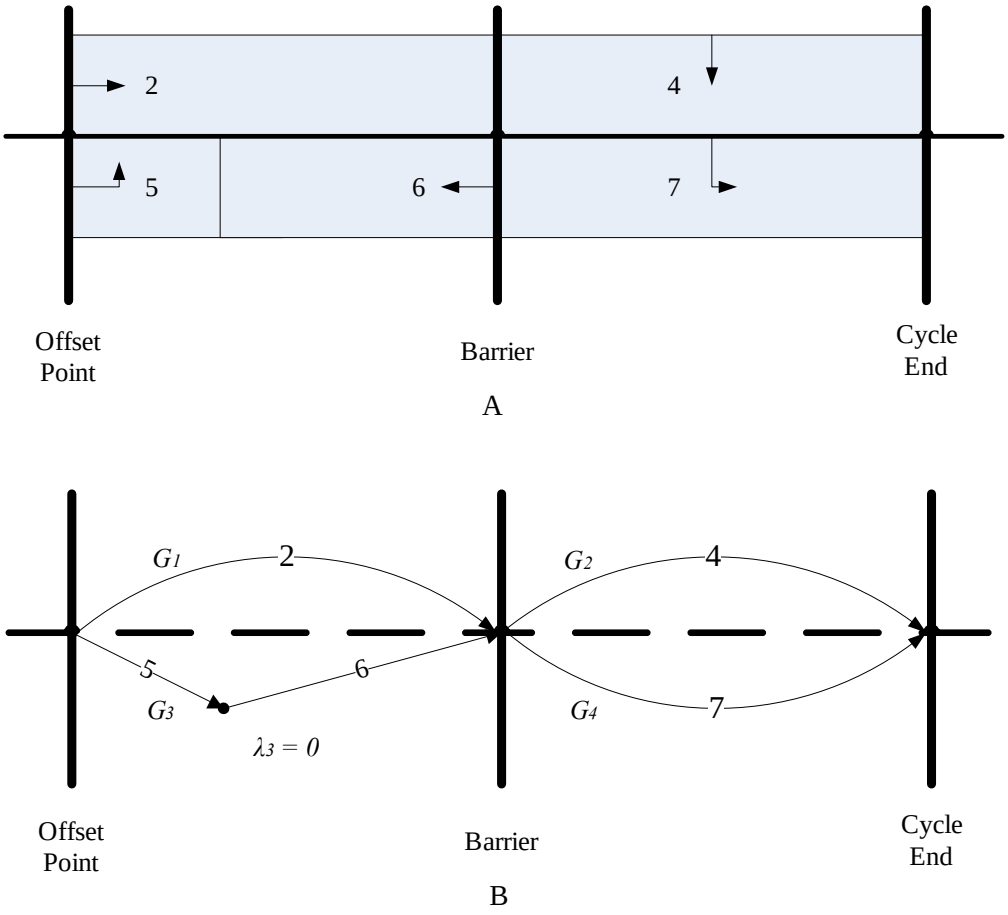


Figure 4-6. NEMA phasing structure for a T-intersection. A) NEMA phasing. B) Transformation of NEMA phasing.

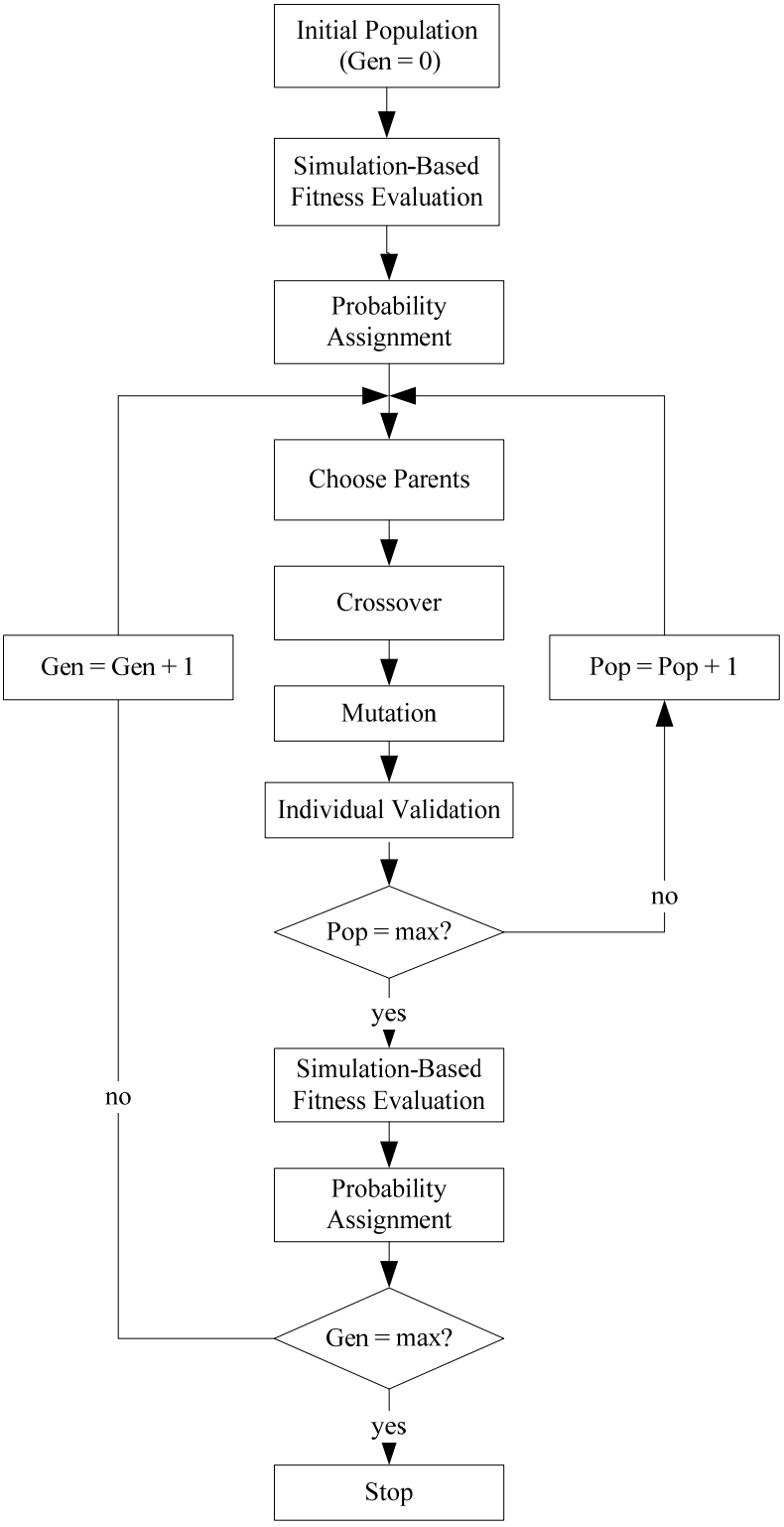


Figure 4-7. Flow chart of the simulation-based GA

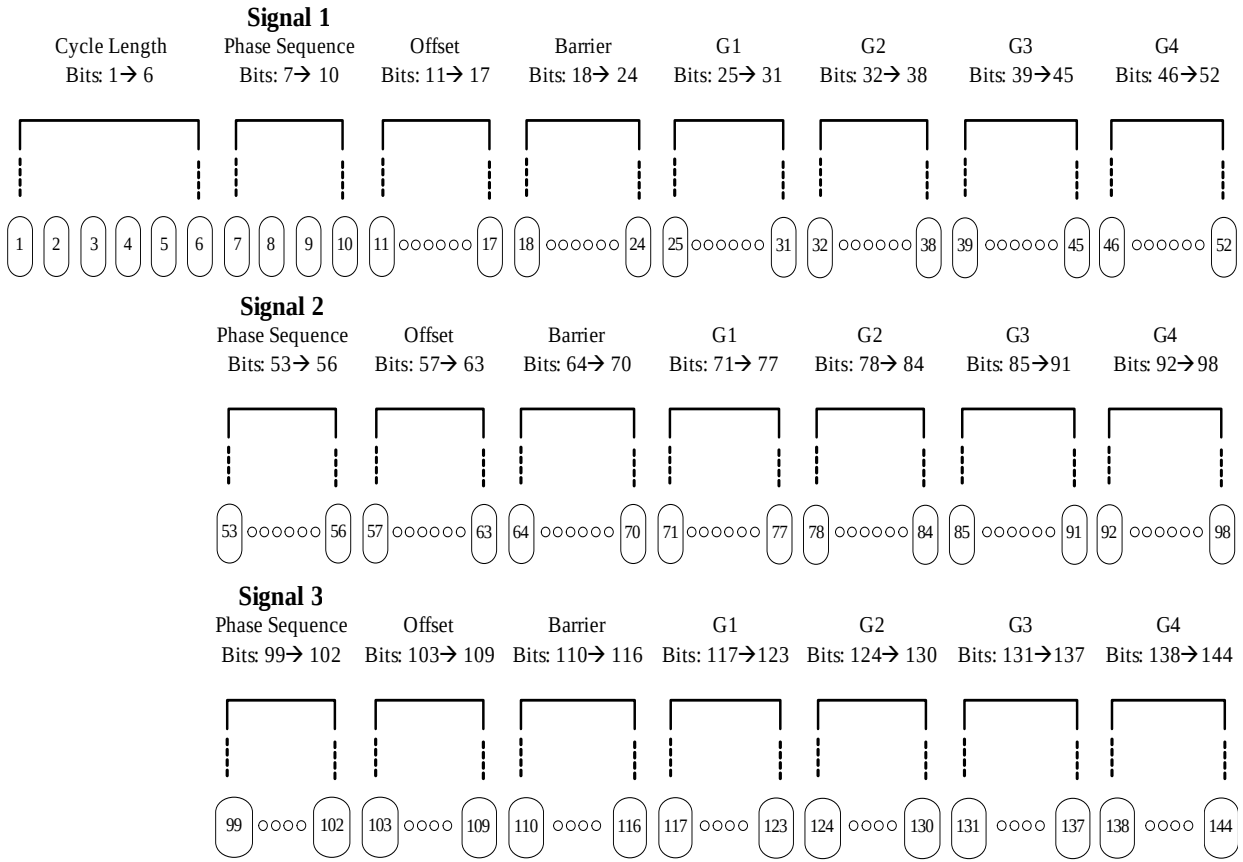


Figure 4-8. Configuration of the cell chromosome

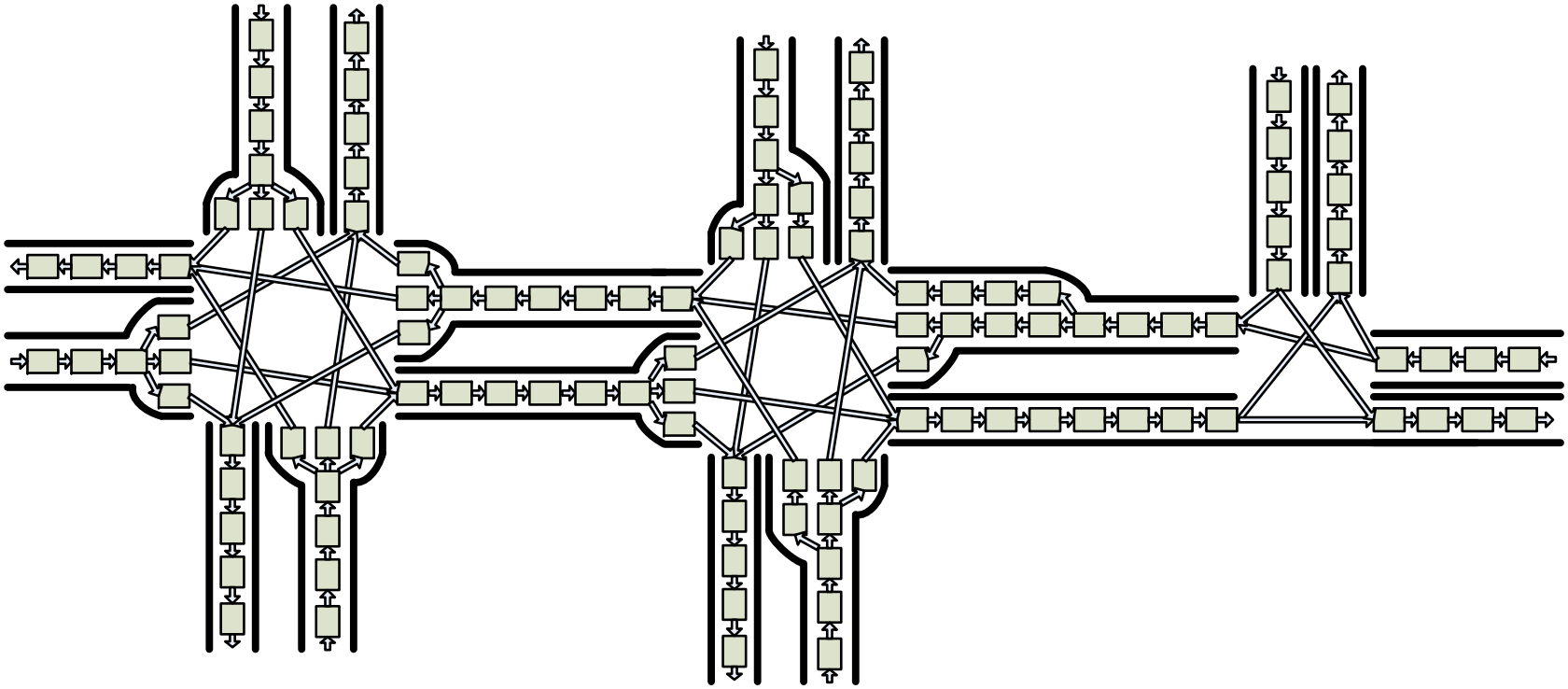
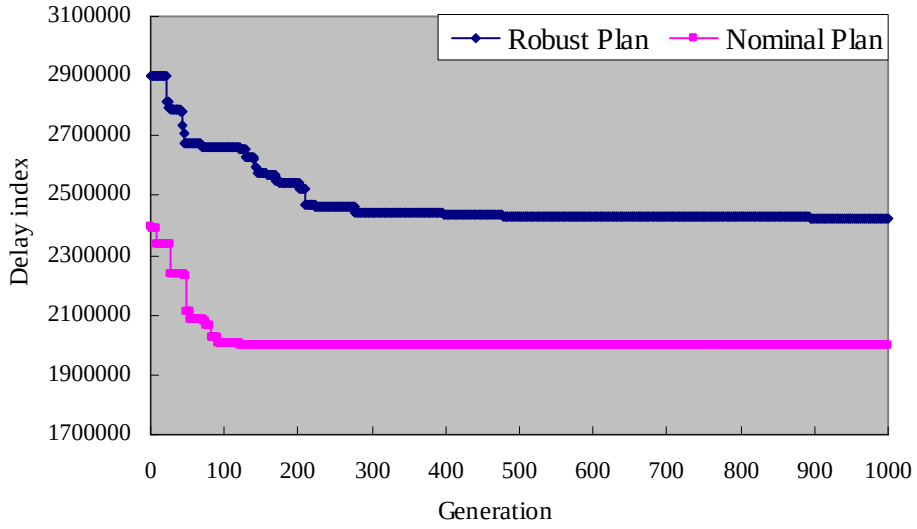
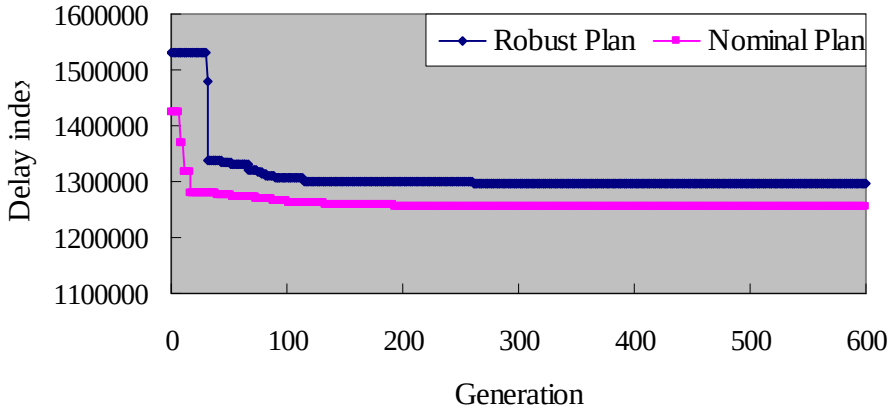


Figure 4-9. Cell representation of the three-node network



A



B

Figure 4-10. Convergence of GA under both traffic conditions. A) Uncongested case. B) Congested Case.

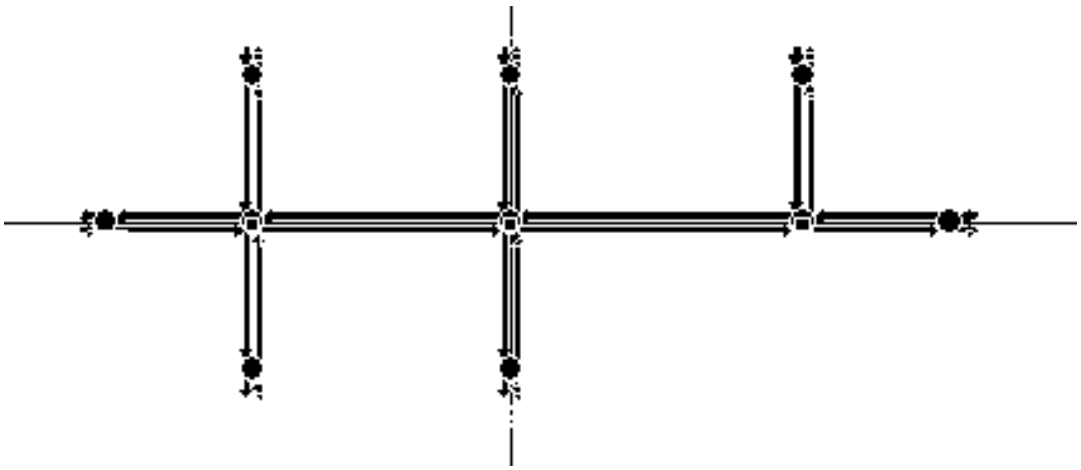


Figure 4-11. A snapshot of the three-node network in CORSIM

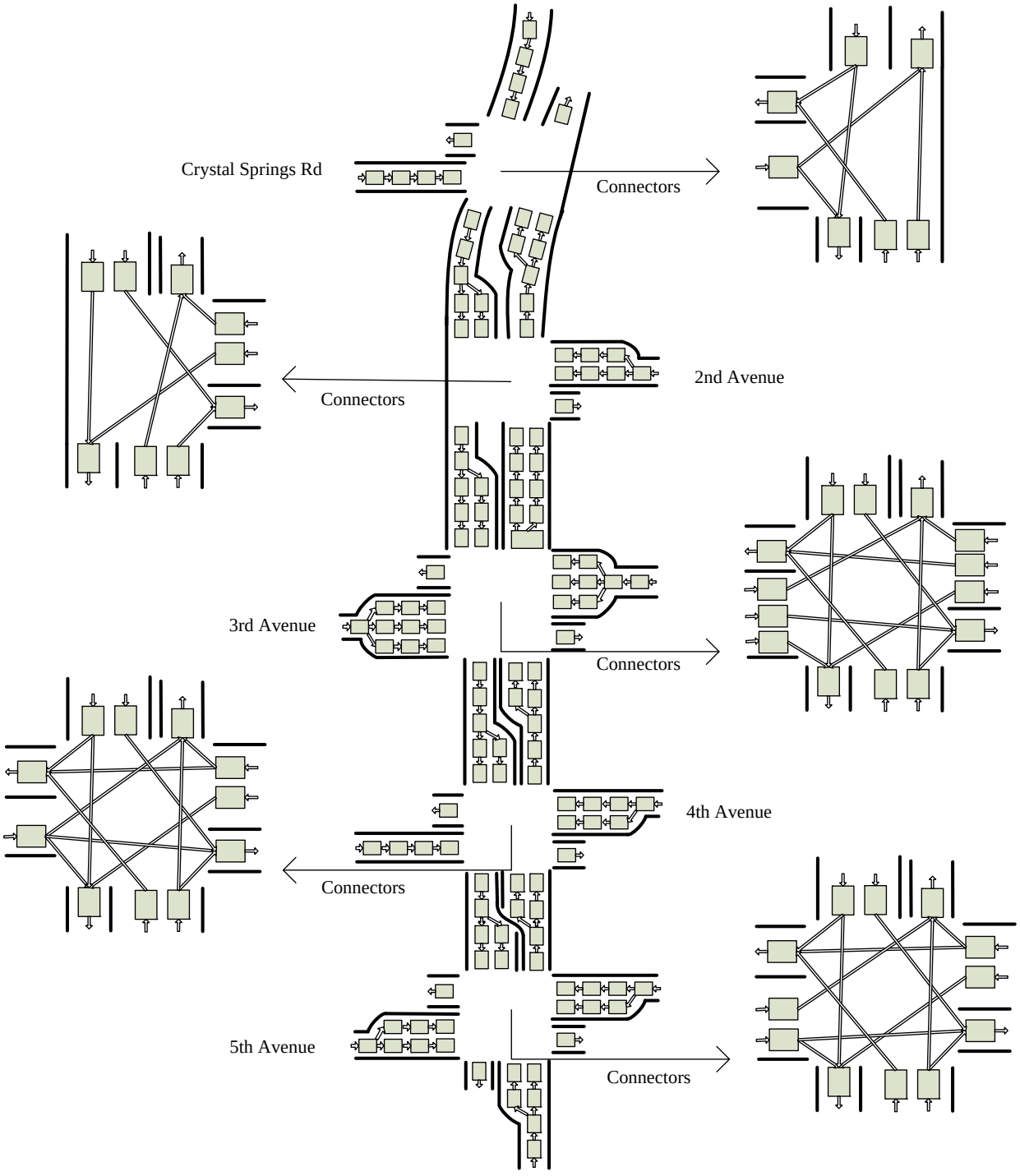
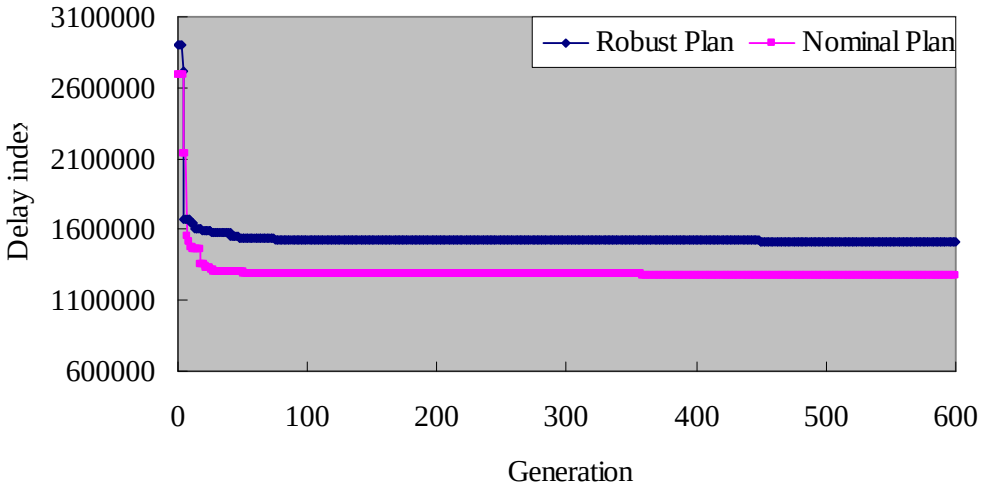
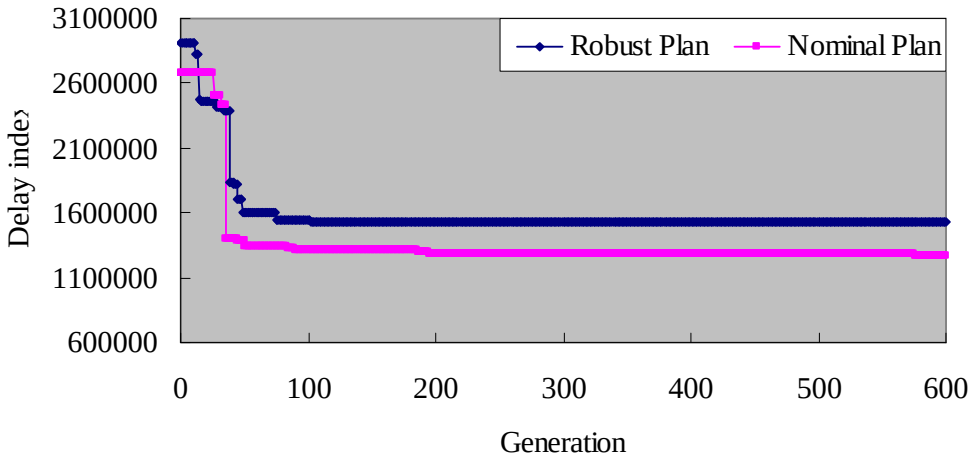


Figure 4-12. Cell representation of El Camino Real arterial



A



B

Figure 4-13. Convergence of GA with different fitness functions. A) 5th-form fitness function. B) Log-form fitness function.

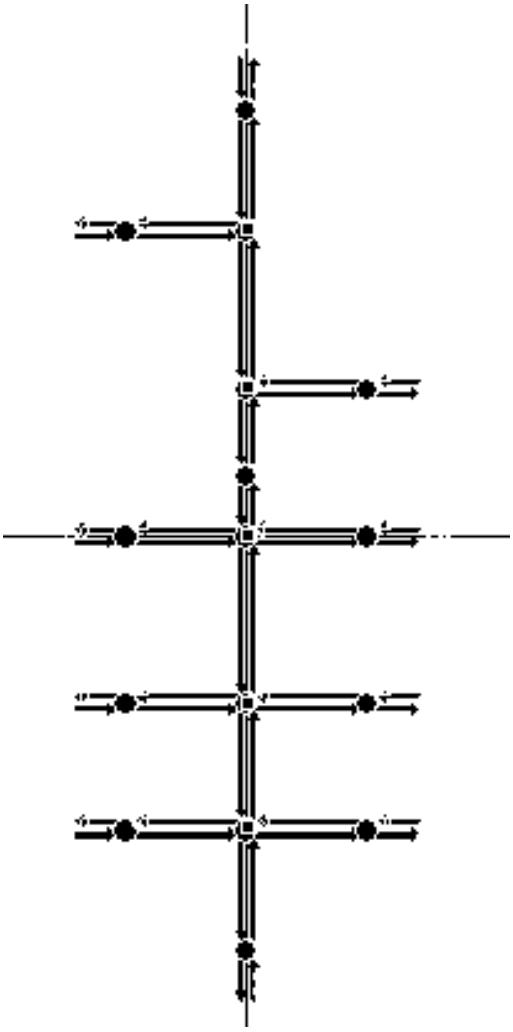


Figure 4-14. A snapshot of the El Camino Real arterial in CORSIM

Table 4-1. CPF result under different 0-1 combinations

0-1combination (χ_1, χ_2)	Constraint representations			
(0, 0)	$y_i(t) = n_i(t)$	$y_i(t) \le Q_{i,\max}(t)$	$y_i(t) \le Q_{i+1,\max}(t)$	$y_i(t) \le \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$
(0, 1)	$y_i(t) \le n_i(t)$	$y_i(t) = Q_{i,\max}(t)$	$y_i(t) \le Q_{i+1,\max}(t)$	$y_i(t) \le \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$
(1,0)	$y_i(t) \le n_i(t)$	$y_i(t) \le Q_{i,\max}(t)$	$y_i(t) = Q_{i+1,\max}(t)$	$y_i(t) \le \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$
(1,1)	$y_i(t) \le n_i(t)$	$y_i(t) \le Q_{i,\max}(t)$	$y_i(t) \le Q_{i+1,\max}(t)$	$y_i(t) = \omega \cdot [N_{i+1,\max} - n_{i+1}(t)]$

Table 4-2. Traffic data for three-node network

Traffic volume		Westbound	Northbound	Eastbound	Southbound
Signal 1	Low demand	--	**162 ±27***	**1223 ±180**	**125 ±18**
	High demand	--	**462 ±45**	**1523 ±270**	**325 ±36**
	Left	0.1445	0.2876	0.0291	0.6637
	Through	0.5772	0.1373	0.8960	0.2389
	Right	0.2783	0.5752	0.0750	0.0973
Signal 2	Low demand	--	**117 ±18**	--	**169 ±27**
	High demand	--	**317 ±36**	--	**269 ±45**
	Left	0.0061	0.7664	0.0778	0.1923
	Through	0.9703	0.0935	0.7243	0.0577
	Right	0.0237	0.1402	0.1979	0.7500
Signal 3	Low demand	**75 ±18**	--	--	**400 ±45**
	High demand	**1075 ±180**	--	--	**400 ±45**
	Left	--	--	--	0.6000
	Through	0.8000	--	1.0000	--
	Right	0.2000	--	--	0.4000

*: **a ± b** means that demand is uniformly distributed in the interval (a-b, a+b) vehicles per hour.

Table 4-3. Signal plans for three-node network

Uncongested case		Cycle length	Phase sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust plan	S1	80	(1, 0, 1, 1)	0	4	58	4	14	4	58	8	10
	S2	80	(0, 1, 1, 0)	76	6	58	4	12	4	60	6	10
	S3	80	(1, 0, 0, 1)	26	--*	50	--	30	32	18	30	--
Nominal plan	S1	80	(1, 1, 0, 1)	0	4	62	4	10	22	44	4	10
	S2	80	(1, 0, 0, 1)	16	4	46	6	24	6	44	4	26
	S3	80	(0, 1, 0, 0)	28	--	50	--	30	22	28	30	--
Congested case		Cycle length	Phase sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust plan	S1	108	(1, 1, 1, 1)	0	8	78	8	14	18	68	8	14
	S2	108	(0, 0, 0, 0)	98	14	72	8	14	28	58	8	14
	S3	108	(1, 0, 1, 0)	16	--	68	--	40	32	36	40	--
Nominal plan	S1	80	(1, 1, 0, 1)	0	4	66	4	6	20	50	4	6
	S2	80	(1, 0, 1, 0)	10	4	46	4	26	4	46	6	24
	S3	80	(0, 0, 1, 1)	4	--	76	--	4	34	42	4	--

*: phase not applicable.

Table 4-4. CORSIM result for three-node network

Traffic condition	Index measure	Robust plan	Nominal plan	Change
Uncongested case	Mean delay	13.15*	17.70	-25.69%
	Mean excess delay	14.02	19.66	-28.68%
Congested case	Mean delay	79.16	98.97	-20.06%
	Mean excess delay	106.58	115.18	-7.46%

*: in vehicle hours.

Table 4-5. Traffic data for El Camino Real arterial

Traffic volume		Westbound	Northbound	Eastbound	Southbound
Crystal Spring	Demand mean	-- *	--	**179**	**1112**
	Demand SD	--	--	**13**	**44**
	Left	--	0.0603	0.7205	--
	Through	--	0.9397	--	0.8369
	Right	--	--	0.2795	0.1631
2nd Ave	Demand mean	**174**	--	--	--
	Demand SD	**18**	--	--	--
	Left	0.6647	--	--	0.1118
	Through	--	0.7818	--	0.8882
	Right	0.3353	0.2182	--	--
3rd Ave	Demand mean	**270**	**--**	**238**	--
	Demand SD	**22**	**--**	**18**	--
	Left	0.4545	0.0600	0.1911	0.0247
	Through	0.2557	0.8679	0.4837	0.8983
	Right	0.2898	0.0722	0.3252	0.0770
4th Ave	Demand mean	**528**	**--**	**101**	--
	Demand SD	**32**	**--**	**10**	
	Left	0.3351	0.0237	0.1272	0.1193
	Through	0.2990	0.8732	0.6301	0.8593
	Right	0.3660	0.1031	0.2428	0.0214
5th Ave	Demand mean	**219**	**1443**	**184**	--
	Demand SD	**10**	**82**	**18**	--
	Left	0.3853	0.0447	0.2889	0.0312
	Through	0.4391	0.9407	0.5804	0.8982
	Right	0.1756	0.0147	0.1307	0.0706

*: -- means data not applicable or available.

Table 4-6. Signal plans for El Camino Real arterial

5^{th}-form fitness function		Cycle length	Phase sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust plan	S1	90	(0, 1, 1, 1)	0	11	58	8	13	47	22	12	9
	S2	90	(0, 0, 1, 0)	15	16	34	26	14	34	16	9	31
	S3	90	(0, 1, 1, 1)	13	24	33	23	10	38	19	17	16
	S4	90	(0, 0, 0, 0)	67	32	18	40	--	--	50	--	40
	S5	90	(0, 0, 0, 0)	0	--	35	--	55	15	20	55	--
Nominal plan	S1	112	(0, 0, 0, 0)	0	14	76	14	8	78	12	9	13
	S2	112	(1, 0, 0, 1)	34	32	55	17	8	70	17	10	15
	S3	112	(1, 0, 0, 1)	38	26	43	25	18	57	12	24	19
	S4	112	(0, 0, 1, 1)	83	29	21	62	--	--	50	--	62
	S5	112	(1, 1, 0, 0)	43	--	68	--	44	21	47	44	--
Log-form fitness function		Cycle length	Phase sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Robust plan	S1	94	(0, 0, 0, 0)	0	8	63	15	8	62	9	10	13
	S2	94	(0, 0, 1, 1)	24	16	35	35	8	35	16	27	16
	S3	94	(1, 0, 0, 1)	60	21	32	30	11	26	27	28	13
	S4	94	(1, 1, 1, 0)	13	29	10	55	--	--	39	--	55
	S5	94	(1, 0, 0, 0)	69	--	55	--	39	44	11	39	--
Nominal plan	S1	112	(1, 0, 1, 1)	0	8	80	16	8	79	9	14	10
	S2	112	(0, 1, 1, 0)	12	11	60	3	33	45	26	31	10
	S3	112	(1, 1, 1, 0)	26	26	45	13	28	62	9	19	22
	S4	112	(1, 1, 1, 1)	10	32	22	58	--	--	54	--	58
	S5	112	(0, 0, 1, 1)	50	--	27	--	85	19	8	85	--

Table 4-7. CORSIM result for El Camino Real arterial

Fitness function	Index measure	Robust plan	Nominal plan	Change
Log-form	Mean delay	234.36	307.13	-23.69%
	Mean excess delay	240.92	312.08	-22.80%
5^{th}-form	Mean delay	228.02	277.46	-17.82%
	Mean excess delay	232.30	281.02	-17.34%

CHAPTER 5
SIGNAL TIMING OPTIMIZATION WITH ENVIRONMENAL CONCERNS

Previous studies have developed signal timing optimization models to minimize both congestion and emissions. For example, Li et al. (2004) optimized the cycle length and green splits for an isolated intersection to minimize a weighed sum of the delay, fuel consumption and emission. Park et al. (2009) developed a signal timing optimization model to minimize the fuel consumption and vehicle emission based on microscopic simulation. These previous studies consider tailpipe emissions instead of roadside air pollution concentrations, which decide the air quality. In other words, dispersion of air pollutants is not captured in their models. The tradeoff between delays and emissions in street canyons should be very different from that in open rural environments, because air pollutants are difficult to disperse in the former, resulting in higher pollution concentrations and worse air quality in the area adjacent to the streets. Moreover, previous studies either use a highly simplified approach to compute traffic emissions or rely on microscopic driving-cycle-based emission models. The former approach fails to capture spatially and temporally varying traffic state and may underestimate the emissions due to accelerations and decelerations (Lin and Ge, 2006). The latter approach results in microscopic-simulation-based optimization models that are computationally intractable for real network optimization.

In contrast, this chapter tries to apply the scenario-based stochastic programming approach to determine traffic signal timings to minimize traffic delay and the risk associated with human exposure to traffic emissions. The model is macroscopic and computationally tractable. At the

same time, it is sufficiently accurate in modeling traffic dynamics and capturing the impact of time-dependent traffic characteristics on emissions.

5.1 Emission and Roadside Pollutant Concentrations

5.1.1 Emission Model

Existing emission modeling systems can estimate and predict traffic emission at regional, facility and individual-vehicle levels. See, e.g., Yu et al. (2010), for a recent review on the emission models. For the analysis of signalized intersection emissions, the modal and instantaneous emission models are more applicable because they predict second-by-second tailpipe emissions as a function of the vehicle's operating mode (e.g., Cernuschi et al., 1995; Barth et al., 2000 and Rakha et al., 2004).

In this research, emission factors are estimated using the modal emission approach proposed by Frey et al. (2001, 2002) where a vehicle's operating mode is divided into four modes of idle, acceleration and deceleration and cruise, and the average emission factors for each mode are determined. As demonstrated below, the resolution and data requirement of this approach are very compatible with the CTM representation of traffic dynamics. Moreover, it is efficient enough to be incorporated into the signal timing optimization procedure.

Frey et al. (2001) defined the idle mode as zero speed and zero acceleration. For the acceleration mode, the vehicle speed must be greater than zero with an acceleration rate averaging at least 1 mph/s for 3 seconds or more. Deceleration is defined in a similar manner as acceleration but with a negative acceleration rate. The cruising mode is approximately steady-speed driving, but some drifting of speed is allowed. In the CTM implementation, the traffic state

at each cell is determined at each time interval and the driving mode of the vehicles in the cell[1] can then be identified through comparing the densities of the current and the immediately downstream cells.

Figure 5-1 illustrates how the driving mode of an ordinary cell is identified. The current cell is *i* and the downstream cell is *i* +1 and their densities are denoted as k_i and k_{i+1}. The fundamental diagram of the signalized arterial is also presented in the figure to facilitate the presentation, where k_{jam} is the jam density and k_c is the critical density. In Figure 5-1, the relationship between k_i and k_{i+1} falls into one of the four areas, i.e., C, A, D and I. Area "C" stands for the cruise mode and consists of two portions: the first portion is a rectangle for the situations when $k_i \le k_c$ and $k_{i+1} \le k_c$, i.e., the vehicles in both cells are traveling at the free-flow speed; the other is the line separating the areas of "A" and "D", where $k_i = k_{i+1} < k_{jam}$, implying that vehicles in both cells are traveling at the same positive speed. The end of that line is the area (point) of "I", representing the idle mode. At this point, $k_i = k_{i+1} = k_{jam}$, and all vehicles have a zero speed. Albeit a point, it occurs frequently in the cells near signalized intersections during the red intervals. The lower area of "D" represents the deceleration mode where $k_i < k_{i+1}$, and $k_{i+1} > k_c$. This implies that vehicles in the current cell are traveling faster than the downstream vehicles. Because these vehicles will move into the downstream cell, they have to decelerate to adapt to the prevailing downstream speed. Similarly, Area "A" represents the acceleration mode where $k_i > k_{i+1}$ and $k_i > k_c$, implying that vehicles in the current cell are traveling slower than

[1] In CTM, vehicles in one cell are treated homogenous.

the downstream vehicles, and are accelerating. In summary, by simply comparing the cell density with that of the immediately downstream cell, one can easily identify the driving mode of the vehicles in each cell. As an example, Figure 5-1 shows a case that the vehicles in cell *i* are in the deceleration mode.

Figure 5-1 describes the determination of driving mode of an ordinary cell. Necessary amendments are introduced to determine the driving modes for other type of cells. For example, the driving mode of signalized cells will be idle when the corresponding signal phase is red, irrelevant of the density of the downstream cell.

5.1.2 Pollution Dispersion Model

Frey et al. (2001, 2002) developed the average emission factors for each driving mode. Applying these factors, the cell emission rate can be calculated as follows:

$$ER_i(t) = n_i(t) \times EF_i(t)$$

where $ER_i(t)$ denotes the emission rate of cell *i* at time interval *t*; $n_i(t)$ is the number of vehicles in the cell; $EF_i(t)$ is the emission factor for a driving mode in units of mg/s or g/s.

With the cell emission rates, an atmospheric dispersion model can be used to estimate the roadside air pollutant concentrations. To facilitate the presentation of the modeling framework, this research adopts the simplest steady-state Gaussian plume dispersion model (e.g., Turner, 1994). Other more advanced dispersion models can be applied in a similar way, particularly those non-steady-state models to capture the time-varying emission rates computed above. However, we note that the dispersion model should be computationally tractable so that it can be incorporated into the signal timing optimization procedure.

To apply the Gaussian plume dispersion model, it is assumed that the pollution source of each cell is located in the cell center. Within the modeling horizon, e.g., 1 hour, the average cell emission rate ER_i is estimated as $ER_i = \sum_{t=1}^{T} ER_i(t)/T$. Subsequently, at any point in space (x, y, z), the incremental pollutant concentration $C_i(x, y, z)$ from the source cell i can be approximately estimated as follows:

$$C_i(x,y,z) = \frac{ER_i}{2\pi U \sigma_y \sigma_z} \exp\left(\frac{-y^2}{2\sigma_y^2}\right)\left\{\exp\left[\frac{-(z-h)^2}{2\sigma_z^2}\right] + \exp\left[\frac{-(z+h)^2}{2\sigma_z^2}\right]\right\}$$

where U is the wind speed at the pollutant release height h; x is the distance along the wind direction; y is the distance along the cross-wind direction; z is the vertical distance; σ_y is the cross-wind dispersion coefficient and σ_z is the vertical dispersion coefficient.

For numerical computation, the roadside area is cut into grids and the pollutant concentrations of each grid are computed with the coordinates of the center of the grid.

Figure 5-2 presents two concentration contour maps for one particular signalized arterial. The maps are generated using the same wind speed but different directions. The wind direction in Figure 5-2A is $\pi/4$ clockwise to the north while in Figure 5-2B it is the north. These two examples demonstrate the need for considering pollutant dispersion as the impacts of traffic emissions will be substantially different in these two cases.

5.1.3 Mean Excess Exposure

We define a surrogate measure to represent the potential negative impacts of roadside pollutant concentrations. The so-called total human emission exposure is calculated as follows:

$$TE = \sum_{i=1}^{I} \int_{z} \int_{y} \int_{x} p(x,y,z) \cdot C_i(x,y,z) dx dy dz \tag{5-1}$$

where TE is the total human emission exposure; I is the total number of cells of the signalized corridor and $p(x,y,z)$ is the population density function beside the corridor.

The above definition is written for one particular pollutant for illustration purposes. If multiple pollutants are considered, Equation (5-1) can be applied to each pollutant and the total human emission exposure is the weighted sum of the exposures to all pollutants. The weighting factors are determined based on, e.g., the social cost of each pollutant (Delucchi, et al., 2002).

The above human emission exposure is computed for a certain wind speed U and direction WD. These two parameters change constantly while the signal plan implemented in the field remains the same for a certain period of time, e.g., years. Therefore, the changing wind speed and direction should be proactively considered. To represent such wind uncertainty, we introduce a set of wind scenarios $\Omega = \{1,2,3,\Lambda\ ,S\}$. A scenario consists of a certain wind speed and direction, i.e., $\omega^s = (U^s, WD^s)$, and its probability of occurrence is p^s. The wind scenarios can be specified using the wind rose of the study area. Figure 5-3 is the 30-year wind rose map for the month of July obtained from a wind station in San Francisco bay area.

Given the set of wind scenarios, it is straightforward to estimate the mean of the human emission exposures across all scenarios, and then optimize signal timings to minimize it. However, decision makers may be more concerned with worst-case scenarios where substantial emission exposure may occur. To address such a risk-averse attitude and avoid being too conservative at the same time, we optimize the signal timings against a set of worst-case

scenarios. More specifically, we minimize the expected emission exposure incurred by those high-consequence scenarios whose collective probability of occurrence is $1-\alpha$, where α is a specified confidence level, e.g., 80%. In financial engineering, the performance measure is known as conditional value-at-risk (CVAR) (Rockafellar and Uryasev, 2000) and we name it as mean excess (emission) exposure.

Mathematically, for a wind scenario ω^s, the total emission exposure can be computed, denoted as TE_s. Consider all wind scenarios and order the emission exposure as $TE_1 < TE_2 < ... < TE_S$. Let s_α be the unique index such that:

$$\sum_{s=1}^{s_\alpha} p_s \geq \alpha > \sum_{s=1}^{s_\alpha - 1} p_s$$

In words, TE_{s_α} is the maximum emission exposure that is exceeded only with probability $1-\alpha$. Consequently, the expected emission exposure exceeding TE_{s_α}, i.e., the mean excess exposure, can be computed as follows:

$$MEE_\alpha = \frac{1}{1-\alpha}\left[\left(\sum_{s=1}^{s_\alpha} p_s - \alpha\right) TE_{s_\alpha} + \sum_{s=s_\alpha+1}^{S} p_s TE_s\right] \quad (5\text{-}2)$$

The second component in the bracket is simply to compute the mean value, and the first is to split the probability 'atom' at the emission exposure point TE_{s_α} to make the collective probability of scenarios considered in the bracket exactly equal to $1-\alpha$. Rockafellar and Uryasev (2002) showed that minimizing Equation (5-2) is equivalent to minimizing the following equation:

$$\min_{\zeta} Z_{\alpha} = \zeta + \frac{1}{1-\alpha}\sum_{s=1}^{S} p_s \cdot \max(TE_s - \zeta, 0) \tag{5-3}$$

where ζ is a free decision variable.

5.2 Model Formulation and Solution Algorithm

5.2.1 Bi-Objective Optimization Model

Given a particular network, the cell representation should be first constructed based on the geometry and signal setting. The cells are then classified into six categories and the corresponding constraints for flow propagation can be written for each cell as previously described in Chapter 4. The constraints comprise a linear system with integer variables. Other constraints include those for signal timings such as minimum and maximum green times, and relationships of phase sequences. For numerical computation of roadside pollutant concentrations, the roadside area is cut into grids and the constraints for computing pollutant concentrations can be written. All these constraints are essentially linear (with respect to decision variables) but involve integer variables. With the linear objective functions to minimize total system delay as described in Figure 4-3 (denoted as TD in this chapter) and the mean excess exposure (Equation (4-3)), the optimization problem is a bi-objective mixed-integer linear program. One portion of the optimal solution to the program specifies the signal timing, denoted as a vector $\left(l^*, o^*, \lambda^*, g^*\right)^T$, where l^*, o^*, λ^* and g^* are vectors of optimal cycle length, offsets, phase sequences and green splits.

Conceptually, the bi-objective signal optimization model is summarized as follows:

$$\min_{(l,o,\lambda,g)} \left\{TD; MEE_{\alpha}\right\}$$

s.t. Constraints for traffic dynamics and relating signal timing to flow propagation

Constraints for feasible signal settings

Constraints for computing pollutant concentrations and human emission exposure

There may not exist an unambiguous optimal solution that minimizes both the total delay and mean excess exposure simultaneously. Hence, a set of Pareto optimal solutions or non-dominated solutions are sought instead. Those solutions are optimal in the sense that no improvement can be achieved in any objective without degradation in the other. All these solutions form a Pareto frontier. Based on the decision maker's consideration of social costs of delays and emissions, an optimal timing plan can be selected.

5.2.2 Simulation-Based Bi-Objective Genetic Algorithm

The bi-objective model formulated above contains a large number of binary variables, particularly when the optimization horizon is long and the network size is large. A genetic-algorithm-based solution approach (Yin, 2002) is adopted to solve the problem for a set of Pareto optimal solutions. The proposed GA follows the general framework of GA. For each individual signal plan, we run a macroscopic simulation based on CTM with all wind scenarios and calculate the corresponding system delays and mean excess exposure. The fitness value of each signal plan is then determined using the modified distance method proposed by Osyczka and Kundu (1996). The basic idea is that if a newly generated Pareto optimal plan is farther away from the existing Pareto set, it will have a greater finiteness value. The detailed algorithm is listed as follows:

Step 1. Let $\gamma = 1$ and $j = 1$, where γ is the generation index and j is the individual index within each generation. Generate the first generation randomly and make the first solution as a starting Pareto solution with an arbitrarily assigned value ρ_1, called the starting latent potential.

Step 2. For each generated solution X, calculate the relative distances from all existing Pareto solutions using the following formula:

$$d_l(X) = \sqrt{\sum_{i=1}^{I}\left(\frac{f_i^l - f_i(X)}{f_i^l}\right)^2} \quad \text{for } l = 1, 2, \ldots, l_{pr}$$

Where l_{pr} is the number of existing Pareto optimal solutions and f_i is the i^{th} objective function in the bi-objective optimization model.

Step 3. Find the minimum value from the set $\{d_l(X)\}$, using the formula:

$$d_{l^*} = \min_{l=1,2\ldots l_p} \{d_l(X)\}$$

Where l^* is the nearest Pareto optimal point to solution X.

Step 4. Compare the newly generated solution X with the existing Pareto solutions in the following way:

- If it is a new Pareto solution, dominating some of the existing Pareto solutions, then calculate its fitness value using $F = \rho_{\max} + d_{l^*}$, and let $\rho_{\max} = F$. Update the Pareto set by removing the solutions dominated by the newly generated solution, then go to Step 5.
- If it is a Pareto solution, dominating none of the existing Pareto solutions, then calculate the fitness value using $F = \rho_{l^*} + d_{l^*}$, and add this solution to the Pareto set with the latent potential value F. If $F > \rho_{\max}$, let $\rho_{\max} = F$. Go to Step 5.
- If it is not a Pareto solution, then update the fitness value as $F = \max(\rho_{l^*} - d_{l^*}, 0)$. Go to Step 5.

Step 5. Substitute $\rho_{\max}$ for all existing Pareto solutions, i.e.

$$\rho_l = \rho_{\max}, \text{ for } l = 1, 2, \ldots, l_p$$

Step 6. Let $j = j + 1$. If $j \le J$, where J is the pre-determined population size. go to Step 2, otherwise go to Step 7.

Step 7. Let $\gamma = \gamma + 1$. If $\gamma \le R$, where R is the pre-determined generation number, perform GA operators: crossover and mutation to generate a new population with size of J. Otherwise, terminate the iterations.

5.3 Numerical Example

The numerical experiment was also carried out on the stretch of El Camino Real described in Chapter 4, between Crystal Springs Rd and 5th Ave. Figure 4-12 is the cell representation of the arterial. The speed limit remains the same, thus the cell length for the major and side streets is 50 and 36 feet respectively. Without considering the demand variations, the mean values in Table 4-5 were used as input demand. The population distribution beside the arterial was assumed to be the following:

$$p(x, y, z) = 3.8 \cdot 10^{-11} \cdot (100 - z) \cdot (30000 - |x|)$$

where $z \le 100$, and $-30000 \le x \le 30000$. In our coordinate system, the y axis is located along the arterial centerline. The above implies that the population density is less for a location farther away from the arterial and the ground level, i.e., with a larger x and z.

To facilitate the presentation, we considered one pollutant only in this example. Table 5-1 presents the emission factors of Nitric Oxide under different driving modes (Frey et al., 2002). Using the 30-year historical wind data from USDA, we specified the wind scenarios. Table 5-2 reports the probabilities of occurrence for all the scenarios we considered. The wind speed is categorized into five bins while the wind direction is captured every $\pi/8$. In total, there are 80 scenarios, but only 37 of them have positive probability of occurrence.

The simulation-based GA was applied to solve the bi-objective signal timing optimization model. The resulting Pareto frontier is shown in Figure 5-4 where one point represents a particular signal timing plan. For example, Plan A minimizes the mean excess exposure and Plan B minimizes the total system delay. The plans in between are all other Pareto optimal solutions, not dominated by any other timing plan. It can be observed from Figure 5-4 that the mean excess exposure resulting from these timing plans varies from 5108 to 2585 person-grams, a 49% reduction. Correspondingly, the total system delay changes from 1325 to 1949 vehicle-hours, a 47% increase. The frontier presents the tradeoff between congestion and emissions, allowing decision makers to learn about the problem before committing to a final decision of an optimal timing plan.

As an example, Table 5-3 presents the optimized timing plans A and B where signal S1 is the one at 5th Avenue and S5 at Crystal Spring. To demonstrate the impacts of these two substantially different timing plan, Figures 5-5 and 5-6 present their corresponding emission exposure contour maps. The former is for the average emission exposure across all the possible scenarios at different height levels while the latter focuses on the 20% high-consequence scenarios. Although the contour maps for Plans A and B show similar shapes, the emission exposure under Plan B is slightly larger than that of Plan A. For side streets, the emission exposure under Plan B is visibly larger than that under Plan A.

5.4 Summary

This chapter has formulated a bi-objective traffic signal optimization model to make an explicit tradeoff between delays and roadside air pollution concentrations. The modal emission approach is integrated with the cell transmission model to compute the pollutant emission rates at each time step. A cell-based Gaussian plume dispersion model is encapsulated to capture the pollutant dispersion process to estimate the human emission exposure. A risk measure, namely, the mean excess exposure, is then defined to represent the mean of the human emission exposures against high-consequence wind scenarios. The signal timing optimization optimizes the cycle length, green splits, offset points and phase sequences to minimize the total system delay and the mean excess exposure simultaneously.

A GA based algorithm is employed to solve the bi-level mixed-integer problem. Both the model and the solution algorithm are demonstrated in a numerical example using field data from a selected real-world arterial network. A set of Pareto optimal signal timing plans are generated that form an efficient frontier. The frontier exhibits an obvious tradeoff between these two objective functions, providing a foundation for a sound decision making.

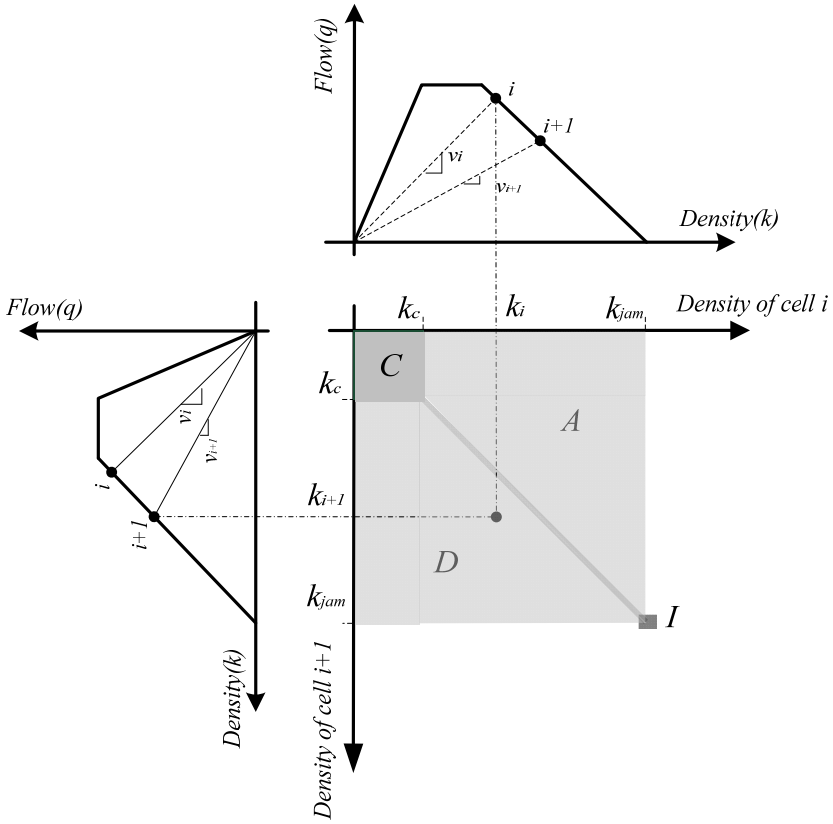
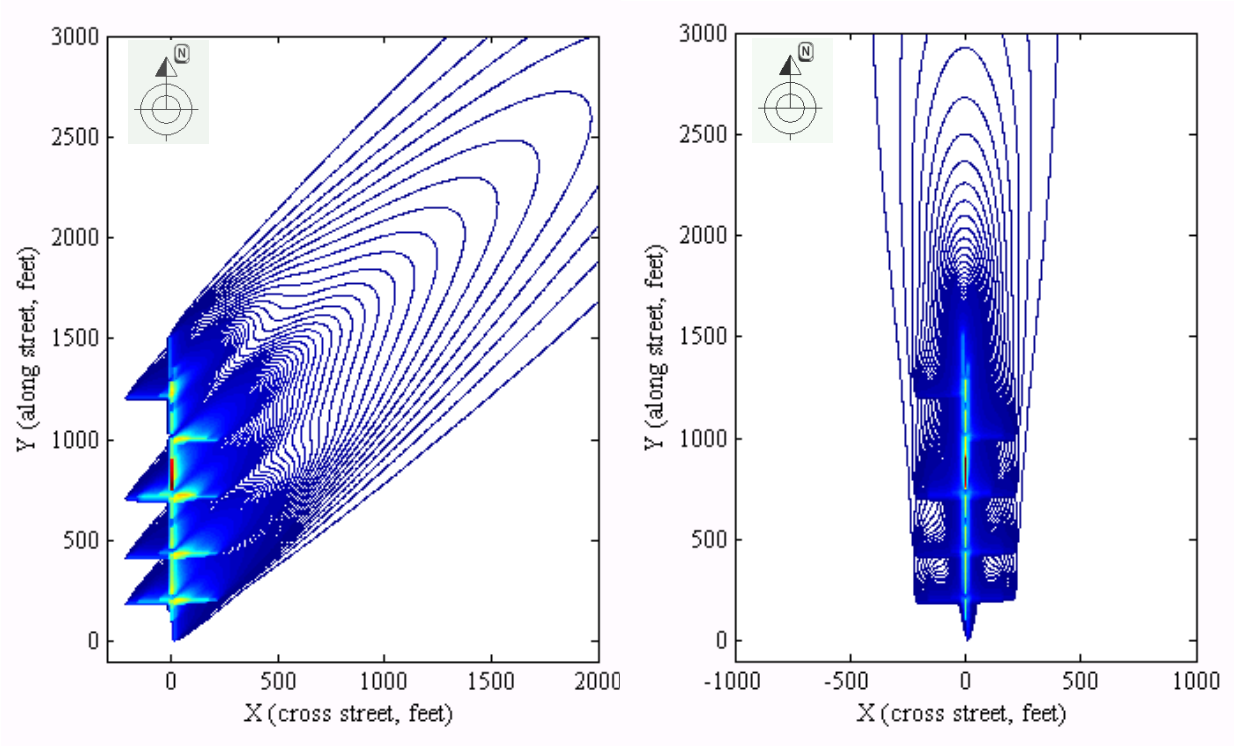


Figure 5-1. Driving mode of cell *i*



A B

Figure 5-2. Examples of pollutant dispersion

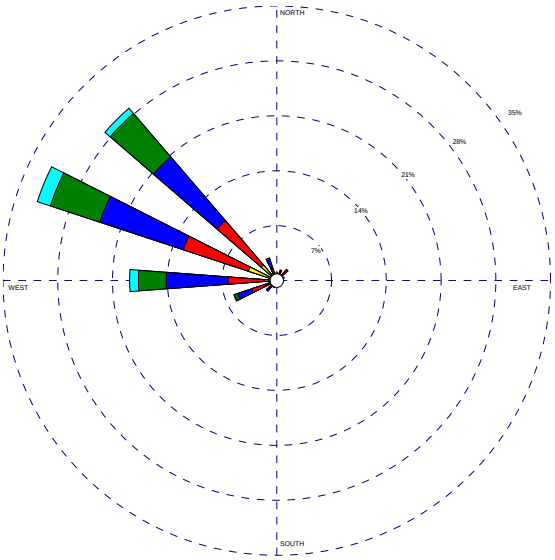


Figure 5-3. An example of wind rose map (source: USDA)

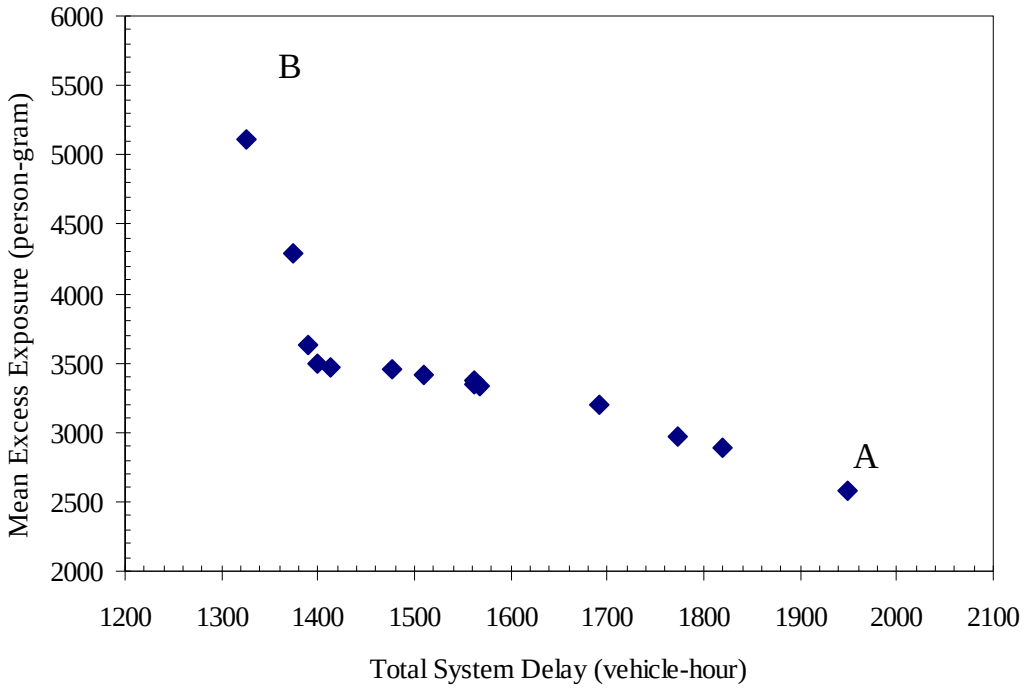
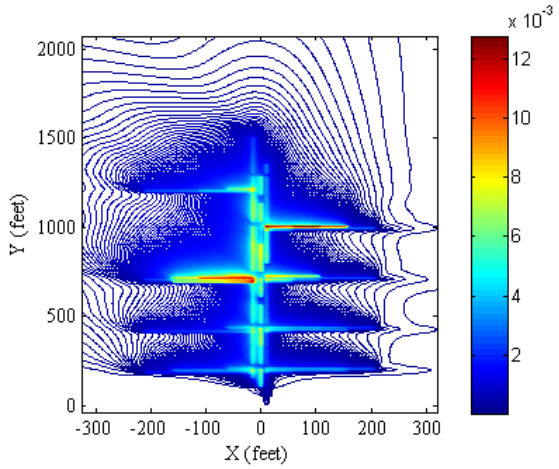
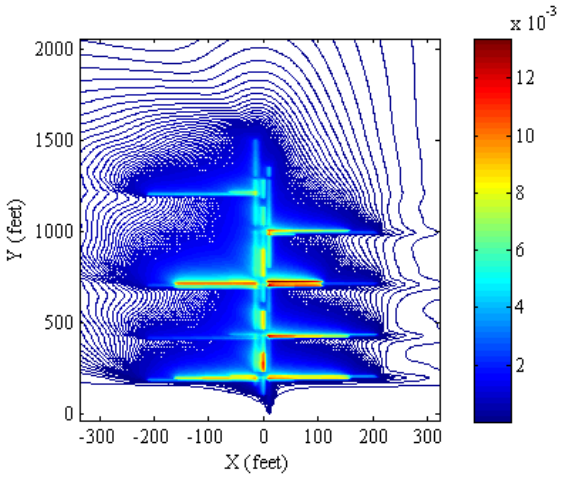


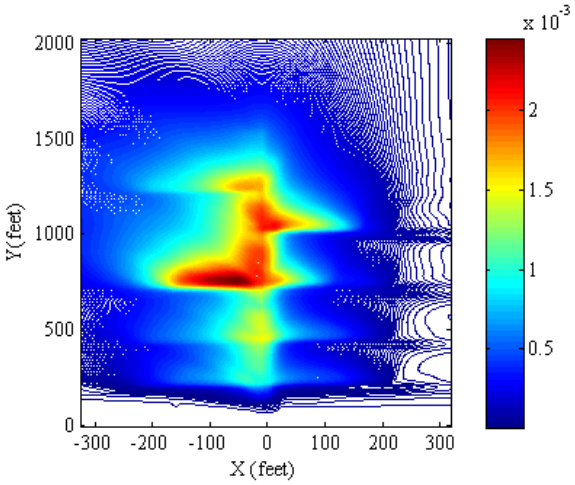
Figure 5-4. Pareto frontier from the GA-based algorithm



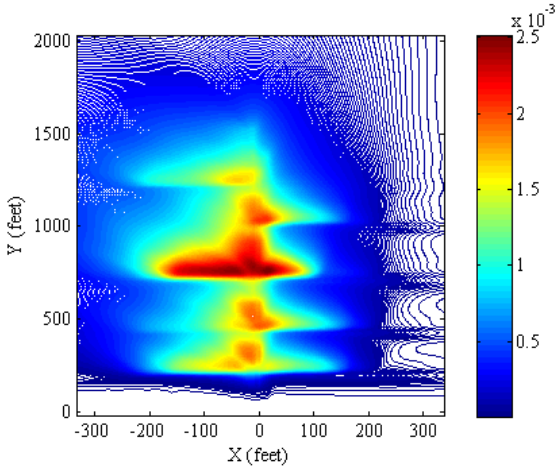
A (Plan A, Z = tailpipe height)



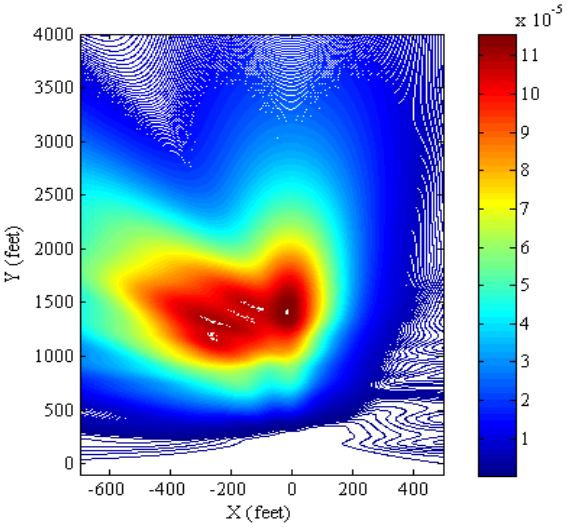
B (Plan B, Z = tailpipe height)



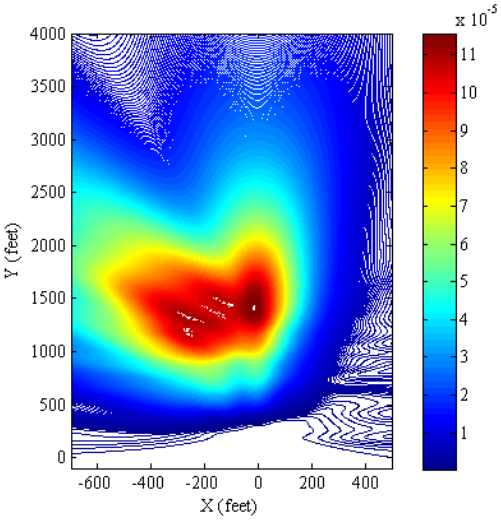
C (Plan A, Z = 6 feet)



D (Plan B, Z = 6 feet)

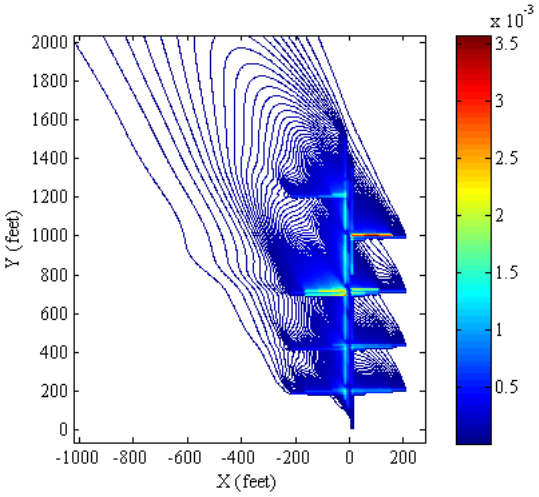


E (Plan A, Z = 50 feet)

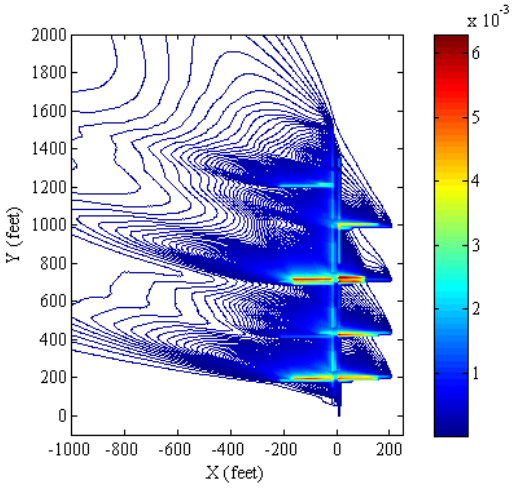


F (Plan B, Z = 50 feet)

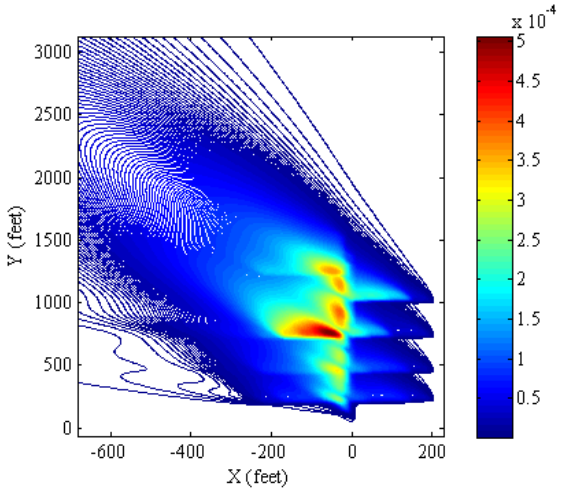
Figure 5-5. Average emission exposure across all scenarios



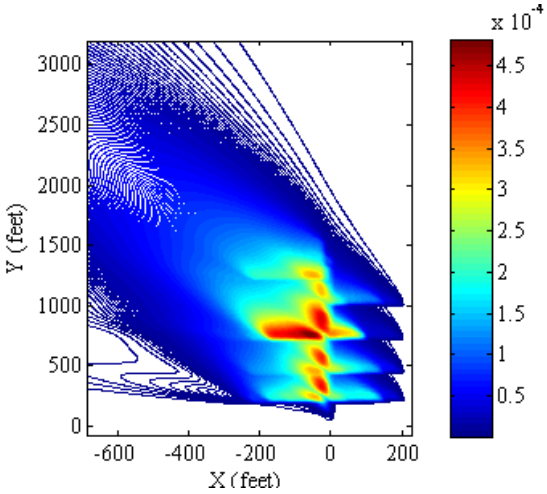
A (Plan A, Z = tailpipe height)



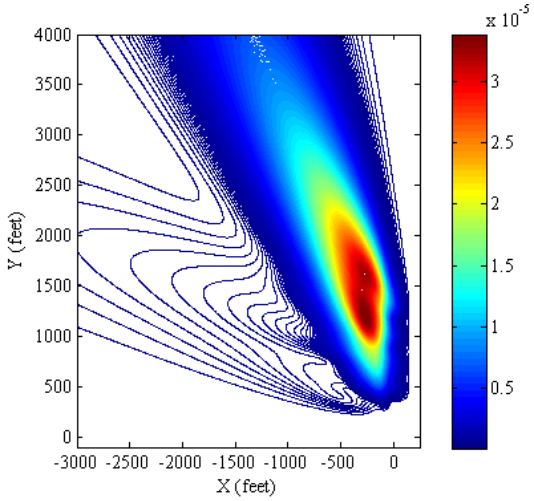
B (Plan B, Z = tailpipe height)



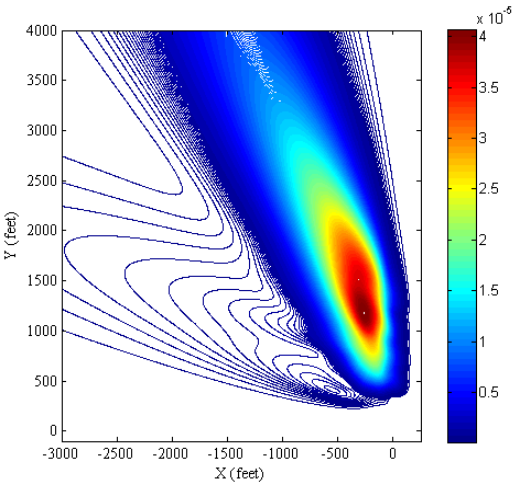
C (Plan A, Z = 6 feet)



D (Plan B, Z = 6 feet)



E (Plan A, Z = 50 feet)



F (Plan B, Z = 50 feet)

Figure 5-6. Average emission exposure across high-consequence scenarios

Table 5-1. Emission factors for different driving modes

Driving mode	Idle (g/s)	Accelerate	Decelerate	Cruise
Emission rate	0.04	0.31	0.06	0.17

Table 5-2. 30-year wind probability data

Wind-speed anti-clock	1.80-3.34 (m/s)	3.34-5.40 (m/s)	5.40-8.49 (m/s)	8.49-11.06 (m/s)	>11.06 (m/s)
0 (north)	0	0.001087	0.000842	0	0
$\pi/8$	0	0.007461	0.013302	0.002029	0.000114
$2\pi/8$	0.018862	0.079837	0.11456	0.078348	0.009455
$3\pi/8$	0.031752	0.094264	0.120749	0.071519	0.01879
$4\pi/8$	0.007061	0.050743	0.085233	0.038261	0.012153
$5\pi/8$	0.002639	0.023677	0.020374	0.004624	0.000883
$6\pi/8$	0	0.004702	0.004019	0.001046	0
$7\pi/8$	0	0	0.000291	0.000428	0
$8\pi/8$	0	0	0	0	0
$9\pi/8$	0	0	0	0	0
$10\pi/8$	0	0	0	0	0
$11\pi/8$	0	0	0	0	0
$12\pi/8$	0	0	0	0	0
$13\pi/8$	0	0.001401	0	0	0
$14\pi/8$	0.001069	0.009813	0.000186	0	0
$15\pi/8$	0	0.005581	0.000294	0	0

Table 5-3. Pareto optimal signal plans under the NEMA phasing structure

Timing plans (secs)		Cycle length	Phase sequence	Offset	P1	P2	P3	P4	P5	P6	P7	P8
Plan A	S1	120	(1, 0, 0, 0)	0	78	16	17	9	18	76	8	18
	S2	120	(1, 0, 1, 0)	38	23	25	61	11	23	25	53	19
	S3	120	(1, 1, 1, 1)	44	24	40	17	56	29	35	48	8
	S4	120	(1, 1, 0, 1)	40	51	47	22	--	--	98	--	22
	S5	120	(1, 0, 1, 0)	40	--	30	--	90	15	15	90	--
Plan B	S1	106	(1, 0, 0, 0)	0	12	66	18	10	45	33	11	17
	S2	106	(0, 1, 0, 1)	52	13	29	10	54	23	19	54	10
	S3	106	(1, 0, 1, 0)	0	28	30	33	15	49	9	25	23
	S4	106	(1, 0, 0, 1)	4	9	11	86	--	--	20	--	86
	S5	106	(0, 1, 1, 0)	14	--	56	--	50	28	28	50	--

CHAPTER 6
ROBUST SYNCHRONIZATION OF ACTUATED SIGNALS ON ARTERIALS

An increasing number of traffic signal controllers used in the United States are traffic-actuated. It has been a common practice to operate these controllers in coordinated systems to provide progression for major traffic movements along arterials and networks. Compared with fixed-time coordinated systems, these semi-actuated coordinated systems offer additional flexibility in responding to fluctuations in traffic demand. Under signal coordination, traffic actuated signals operate under a common background cycle length. Coordination is provided through a fixed reference point, which defines the start of the controller local clock, and can be set to the start of green, end of green (yield point) or other time interval for the sync phases (e.g., beginning of the flashing don't walk interval). Note that the controller local clock definition varies among signal controller manufacturers (FHWA, 1996).

To ensure operation efficiency of coordinated actuated systems, attention should be paid to determining appropriate signal settings, particularly offsets due to the fact that the start of green of the sync phases (typically Phases 2 and 6) is not fixed. This chapter attempts to use the scenario-based approach to synchronize actuated signals along corridors. The model developed is based on Little's mixed-integer linear programming (MILP) formulation (Little, 1966) maximizing the two-way bandwidth to synchronize signals along arterials by determining offsets and progress speed adjustment etc. By specifying scenarios as realizations of uncertain red times of sync phases, we define the regret associated with a coordination plan with respect to each scenario, and then formulated a robust counterpart of Little's formulation as another MILP to minimize the average regret incurred by a set of high-consequence scenarios.

6.1 Bandwidth Maximization for Arterial Signal Coordination

Generally speaking, there are two approaches of generating coordination plans to synchronize signals along arterials and grid networks. One aims at bandwidth maximization, e.g., *MAXBAND* (Little et al., 1981) and *PASSER-II* (Chang et al., 1988) while the other is performance-based optimization, synchronizing signals to minimize the performance measures such as control delay and corridor travel time, e.g., *TRANSY-7F* (Wallace et al., 1998). To facilitate the presentation of our robust approach, this study bases the model development on Little's MILP formulation (Little, 1966). which maximizes the two-way bandwidth to synchronize signals along arterials by determining offsets and progress speed adjustment etc. The model has been proven to be a flexible and robust approach for signal synchronization and actually lays the foundation for *MAXBAND*. It has been later extended to consider variable bandwidth, phase sequencing and grid network synchronization (Gartner and Stamatiadis, 2002; Gartner and Stamatiadis, 2004 and Messer et al., 1987).

Below, we summarize Little's MILP formulation. Given a two-way arterial with an arbitrary number of signals, a common background cycle length, and the split information for each signal, the formulation attempts to synchronize the signals to produce a maximum sum of the inbound and outbound bandwidths. Let S_h and S_i be any pair of adjacent signals, and S_i follows S_h in the outbound direction. Figure 6-1 presents the geometry of the green bands between S_i and S_h. The horizontal lines indicate when the sync phases are red, and the zigzag lines represent the vehicle trajectories.

Notations to be used together with those in Figure 6-1 are introduced as follows:

r_i	red time of signal i on the corridor (cycles)
$b(\bar{b})$	Outbound (inbound) bandwidth (cycles).
$t(h,i)\left[\bar{t}(i,h)\right]$	Outbound (inbound) travel time from signal $h(i)$ to $i(h)$ (cycles)
$\phi(h,i)\left[\bar{\phi}(i,h)\right]$	time from the center of red at S_h to the center of a particular red at S_i. The two reds are chosen so that each is immediately to the left (right) of the same outbound (inbound) green band (cycles)
$w_i(\overline{w_i})$	time from the right (left) side of S_i's red to the green band (cycles)
$m(h,i)$	$\phi(h,i)+\bar{\phi}(h,i)$. According to Figure 6-1, $m(h,i)$ must be integer (cycles)
$T_1,\ T_2$	lower and upper bounds on cycle length (s)
z	signal frequency (cycles/s), the inverse of cycle length
$d(h,i)$	distance from S_h to S_i (m). $d_i = d(i,i+1)$
$v_i(\overline{v_i})$	speed between S_i and S_{i+1} outbound (inbound) (m/s)
$e_i, f_i(\overline{e_i}, \overline{f_i})$	lower and upper bounds on outbound (inbound) speed (m/s)

Note that in Little's formulation, in addition to offsets, cycle length, travel speed, and change in speed between street segments are also decision variables, constrained by upper and lower limits. The bandwidth maximization problem is mathematically written as follows:

$$\max_{(b,\bar{b},z,w,\overline{w},t,\bar{t},m)} B = b + \bar{b}$$

$$s.t. \quad 1/T_1 \le z \le 1/T_2 \tag{6-1}$$

$$w_i + b \le 1 - r_i \qquad \forall i = 1,...,n \tag{6-2}$$

$$\overline{w_i} + \bar{b} \le 1 - r_i \qquad \forall i = 1,...,n \tag{6-3}$$

$$(w_i + \overline{w_i}) - (w_{i+1} + \overline{w}_{i+1}) + (t_i + \bar{t}_i) = m_i - (r_i - r_{i+1}) \qquad \forall i = 1,...,n-1 \tag{6-4}$$

$$m_i = \text{integer} \qquad \forall i = 1,...,n-1 \tag{6-5}$$

$$(d_i / f_i)z \le t_i \le (d_i / e_i)z \qquad \forall i = 1,...,n-1 \tag{6-6}$$

$$(d_i / \overline{f}_i)z \le \overline{t}_i \le (d_i / \overline{e}_i)z \qquad \forall i = 1,...,n-1 \qquad (6\text{-}7)$$

$$(d_i / h_i)z \le (d_i / d_{i+1})t_{i+1} - t_i \le (d_i / g_i)z \qquad \forall i = 1,...,n-2 \qquad (6\text{-}8)$$

$$(d_i / \overline{h}_i)z \le (d_i / d_{i+1})\overline{t}_{i+1} - \overline{t}_i \le (d_i / \overline{g}_i)z \qquad \forall i = 1,...,n-2 \qquad (6\text{-}9)$$

$$b, \overline{b}, w_i, \overline{w}_i \ge 0 \qquad \forall i = 1,...,n$$

In the objective function, $b, \overline{b}, w, \overline{w}, t, \overline{t}$ and m represent vectors whose elements are the scalar decision variables $b, \overline{b}, w, \overline{w}, t, \overline{t}$ and m_i. Equation (6-1) is the constraint on cycle length. Equations (6-2) and (6-3) are the constraints on green bandwidth. Equations (6-4) and (6-5) are the integer constraints due to the fact that $\phi(h,i) + \overline{\phi}(h,i)$ must be integer. Equations (6-6) and (6-7) are the constraints associated with travel speed. Equations (6-8) and (6-9) are the constraints on speed changes between adjacent street segments. All of these constraints and the objective function are linear, and (6-4) and (6-5) are integer, therefore the problem is an MILP, which can be efficiently solved by using, e.g., the branch-and-bound algorithms. From the solutions, offsets can be easily obtained as follows according to the geometry in Figure 6-1:

$$\theta_{i+1} = [w_i - w_{i+1} + t_i + 1/2(r_i - r_{i+1})]_+ \qquad \forall i = 1,...,n-1$$

where θ_{i+1} is the offset of signal *i*+1 with respect to *i*; $[x]_+$ is defined as $x - \text{int}(x)$ and $\text{int}(x)$ represents the largest integer not greater than *x*. Hereafter, we denote the vector of offsets as θ.

6.2 Scenario-Based Approach for Robust Synchronization of Actuated Signals

Little's formulation assumes that the durations of minor phases (red times of sync phases) are deterministic. Such an assumption does not hold for actuated signal control, where the phase durations of minor phases vary between zero (skipped) and the corresponding maximum greens. That is the reason for a so-called "early return to green" problem. To illustrate how prevailing the problem of "early return to green" is, Figure 6-2 depicts the histograms for starts of green of Phase 2 and 6 at two selected intersections, Page Mill and Stanford, along El Camino, Palo Alto,

CA. The data were collected from 11:00 am - 3:00 pm between February 6 and March 7, 2005, for a total of 14 weekdays with more than 1,600 cycles. Page Mill is a critical intersection for the corridor with almost equal amounts of mainline and cross-street traffic. Still, the probabilities of "early return to green" are 61% for Phase 2 and 39% for Phase 6. Stanford has low volume of minor-phase traffic, thus the probabilities are as pretty high as 92% for Phase 2 and 94% for Phase 6 respectively. The histograms confirm the assertion made in the previous studies that the problem of "early return to green" should be recognized and explicitly addressed in actuated signal synchronization. Note that in addition to uncertainty of start of green, end of green is also uncertain, especially under lead-lag phase sequence, due to skip or gap-out of the left-turn phase. Figure 6-3 presents the histograms for green terminations of Phase 2 and 6 at Page Mill and Stanford. It can be seen that compared with starts of green, terminations of green have much narrower spans. Under many circumstances, the termination is the force-off point.

The above empirical data shows that for actuated signals, the red times, r_i, in Little's formulation should follow random distributions with supports between zero and the sum of maximum greens of the conflicting minor phases. To represent the uncertainty, a set of scenarios $\Omega = \{1,2,3,\Lambda\ ,K\}$ in introduced, and each individual scenario is composed of the red time of the sync phases[2] at all the signals, which for example at intersection i, is denoted as r_i^k.

We now define the regret function. For each scenario k, we solve Little's MILP formulation to obtain the maximum two-way bandwidth, denoted as B_k^*. Consequently, for any other feasible coordination plan (offsets θ and cycle frequency z) that may not be optimal for scenario k, the regret or loss can be defined as:

[2] Normally Phases 2 and 6. For simplicity, we assume here that both phases start and end at the same time.

$$L_k = B_k^* - B_k(\theta, z) \qquad \forall k$$

where L_k is the regret or loss of the coordination plan (θ, z) with respect to red time scenario k, and B_k is the two-way bandwidth resulted by (θ, z) under scenario *k*.

Then the robust signal coordination plan can be obtained by minimizing the following equation:

$$\min_{\theta,z,\xi} Z_\alpha = \xi + \frac{1}{1-\alpha}\sum_{k=1}^{K} p_k \cdot \max\left(L_k(\theta, z) - \xi, 0\right)$$

Therefore, the scenario-based robust synchronization model for actuated signals can be written as follows:

$$\min_{(b^k, \bar{b}^k, z, w^k, \bar{w}^k, t^k, \bar{t}^k, m^k, \theta, \zeta, U_k, L_k)} Z_\alpha = \zeta + \frac{1}{1-\alpha}\sum_{k=1}^{K} p_k U_k$$

s.t.

$$U_k \geq L_k - \zeta \qquad \forall k = 1,\ldots,K$$

$$U_k \geq 0 \qquad \forall k$$

$$L_k = B_k^* - (b^k + \bar{b}^k) \qquad \forall k$$

$$1/T_1 \leq z \leq 1/T_2$$

$$w_i^k + b^k \leq 1 - r_i^k \qquad \forall i = 1,\ldots,n,\ \forall k$$

$$\bar{w}_i^k + \bar{b}^k \leq 1 - r_i^k \qquad \forall i = 1,\ldots,n,\ \forall k$$

$$(w_i^k + \bar{w}_i^k) - (w_{i+1}^k + \bar{w}_{i+1}^k) + (t_i^k + \bar{t}_i^k) = m_i^k - (r_i^k - r_{i+1}^k) \qquad \forall i = 1,\ldots,n,\ \forall k$$

$$m_i^k = \text{integer} \qquad \forall i = 1,\ldots,n,\ \forall k$$

$$\theta_{i+1} = w_i^k - w_{i+1}^k + t_{i+1}^k + 1/2(r_i^k - r_{i+1}^k) \qquad \forall i = 1,\ldots,n-1,\ \forall k$$

$$(d_i / f_i)z \leq t_i^k \leq (d_i / e_i)z \qquad \forall i = 1,\ldots,n-1,\ \forall k$$

$$(d_i / \bar{f}_i)z \leq \bar{t}_i^k \leq (d_i / \bar{e}_i)z \qquad \forall i = 1,\ldots,n-1,\ \forall k$$

$$(d_i / h_i)z \leq (d_i / d_{i+1})t_{i+1}^k - t_i^k \leq (d_i / g_i)z \qquad \forall i = 1,\ldots,n-2,\ \forall k$$

$$(d_i / \bar{h}_i)z \leq (d_i / d_{i+1})\bar{t}_{i+1}^k - \bar{t}_i^k \leq (d_i / \bar{g}_i)z \qquad \forall i = 1,\ldots,n-2,\ \forall k$$

$$b^k, \bar{b}^k, w^k, \bar{w}^k \geq 0 \qquad \forall k$$

where U_k is an auxiliary decision variable, equal to $\max(L_k(\theta, z) - \zeta, 0)$. Note that B_k^* is pre-determined by solving Little's formulation for each scenario. As formulated, the problem is

another MILP and can be efficiently solved. As demonstrated in pervious studies (Krokhmal et al., 2002; Chen et al., 2006 and Yin, 2008a), such a formulation can offer computational advantages and allow handling a large number of scenarios. Our numerical example below also shows that the problem can be solved in polynomial time.

6.3 Numerical Example

6.3.1 Plan Generation

We solve the robust synchronization formulation for an arterial with six signals, which is included in CORSIM as one of the tutorial examples, named "ActCtrl Example", shown in Figure 6-4. The outbound signals, starting from signal 1, are located at 0, 314, 554, 759, 1012, 1317m respectively. The timing plan for each isolated signal is determined using the Webster's equation. The red times of the sync phases are assumed to be independently normally distributed with means of 0.27, 0.24, 0.50, 0.35, 0.35, 0.43 cycles respectively and the same standard deviation of 0.05 cycles across all the signals. The scenarios are specified by random sampling and are assumed to have equal probability to occur. The upper and lower limits of the cycle length are 100 s and 45 s. Limits on travel speed for different street segments are set to be the same with the upper limit of 17.9 m/s (40 mph) and the lower limit of 13.4 m/s (30 mph). Changes in reciprocal speeds across all the segments are limited between -0.0121 and $0.0121(\text{m/s})^{-1}$, corresponding to a maximum possible change in speed of ±2.3 m/s (±4.9 mph) at the lower limit of the speed and ±3.9 m/s (±8.8 mph) at the upper limit. The confidence level α is selected to be 0.90.

We use an algebraic modeling system called GAMS (Brooke et al., 1992) and CPLEX solver (CPLEX, 2004) to solve the robust synchronization formulation with the number of scenarios varying from 10 to 250. The computation times (in CPU seconds) and the plan differences are presented in Table 6-1, plotted in Figure 6-5.

The plan difference is defined as $\left\|\theta^K - \theta^{250}\right\|_2 / \left\|\theta^{250}\right\|_2$, where θ^K is the plan generated with K scenarios. Within expectation, the difference tends to decrease as the number of scenarios increases. However, it can be observed that relatively small number of scenarios is already enough to produce similar robust plans.

The reported computation times include the times needed to solve Lttile's formulation for each scenario to obtain B_k^*. Regressing the computational times against the number of scenarios, we obtain the following equation, suggesting that the problem may be solved in polynomial time:

$$\text{Time} = 0.3286*(\text{number of scenario}) - 2.6556 \text{ , } R^2 = 0.9962.$$

For the comparison purpose, we also generate a nominal plan following the procedure of deterministic signal coordination, in which we use the mean red times as our estimates, and then use Little's formulation described in part two of the section to obtain the nominal plan.. Both the robust plan (with 250 scenarios) and the nominal plan are reported in Table 6-2.

6.3.2 Plan Evaluation

To evaluate the performance of both nominal and robust plans, we conduct macroscopic Monte Carlo simulation and microscopic simulation in CORSIM. In the Monte Carlo simulation, 2000 samples of red times are drawn from the same normal distributions previously used to generate scenarios for solving the robust synchronization formulation. With each sample, the bandwidths resulted by both the nominal and robust plans are computed. Consequently, several performance measures, including the average, worst-case (minimum) and 90th percentile minimum bandwidths, and the 90% conditional value-at-risk, are calculated and reported in Table 6-3.

The results indicate that the robust coordination plan performs better against high-consequence scenarios, with the worst-case bandwidth increasing by 23.1%, and the 90th

percentile bandwidth by 23.9%, and the conditional value-at-risk decreasing by 17.3%. At the same time, the average bandwidth also increases by 20%. However, since the robust plan is designed to guard against high-consequence scenarios, an improvement of the average performance is not what we should expect and will not be necessarily obtained.

To further validate the robust synchronization formulation, we evaluate the plans with 2000 samples randomly drawn from independent *uniform* distributions with the minimum and maximum values described in Table 6-4. The resulting performance measures are summarized in Table 6-5. The robust plan still outperforms the nominal plan, with the average, worst-case and 90th percentile bandwidths increasing by 16.7%, 22.1% and 22.5%, and the regret decreasing by 16.5%. This examination suggests that the robust formulation is not overly sensitive to the specification of scenarios and using distorted distributions to generate scenarios may still result in robust timing plans.

We recognize the limitation of the bandwidth-based synchronization that traffic flows and intersection capacities are not considered in the optimization criterion (Gartner, 1991), and thus bandwidth maximization does not necessarily optimize other delay-related performance measures. To examine how the robust plan affects those measures, we conduct a microscopic simulation using CORSIM. The robust and nominal plans are implemented respectively in the semi-actuated corridor of "ActCtrl Example" in CORSIM. We select the control delay, corridor travel time, and vehicle stop ratio as the performance measures where vehicle stop ratio is defined as the total number of stops (when speed is lower than 3 mph) divided by the total number of vehicles served by the corridor within the simulation period. The means and standard deviations calculated from ten simulation runs are reported in Table 6-6. We conduct *t* tests and *F* tests to examine whether those performance measures are statistically different. As shown in

Table 6-6, the three *t* values are all greater than the critical *t* value of 2.552 at the significance level of 1%, suggesting that the corridor has better average performance with the robust plan. The *F* values are all smaller than the critical *F* value of 2.44 at the significance level of 10%, indicating that we can not reject the hypothesis that the variances are the same. It should be pointed out that the intention of the CORSIM simulation is to examine whether the robust plan makes the delay-related performance measures worse off. Although the simulation results actually suggest otherwise, we do not expect to always obtain such improvements, since it is not what the robust plan is designed for. The only conclusion we draw from the CORSIM simulation is that the robust plan seems unlikely to worsen the delay-related performance measures.

6.4 Summary

This chapter presented a robust approach to synchronize actuated signals along arterials. The formulation is an MILP, a class of mathematical programs easy to solve using the state-of-the-art solvers. The computational time only increases polynomially as the number of scenarios increases. The resulting robust coordination plan has been demonstrated in a numerical example to perform better against high-consequence scenarios without losing optimality in average. The approach can be used to either design a new coordination plan for implementation or fine-tune the plan offline after implementation. In the latter case, the specification of scenarios is an easy task with the archived signal status data. One may randomly select 50 to 200 red time realizations from the data and assume equal probability of occurrence. To design a new coordination plan where the distributions of red times are normally unknown, we suggest specifying 50 to 200 scenarios as the points that equally divide the red-time intervals into K+1 segments and assume equal probability of occurrence of $1/K$. The suggestion is based on our observation from the numerical experiments that the robust formulation is not overly sensitive to

the specification of scenarios. Even with biased scenarios, the formulation may still produce meaningful robust plans.

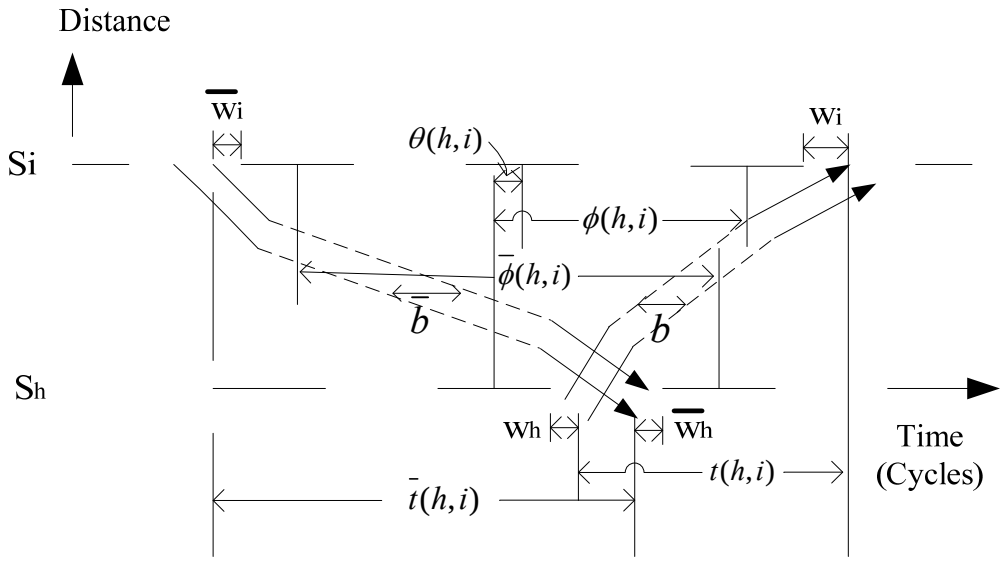


Figure 6-1. Geometry of green bands

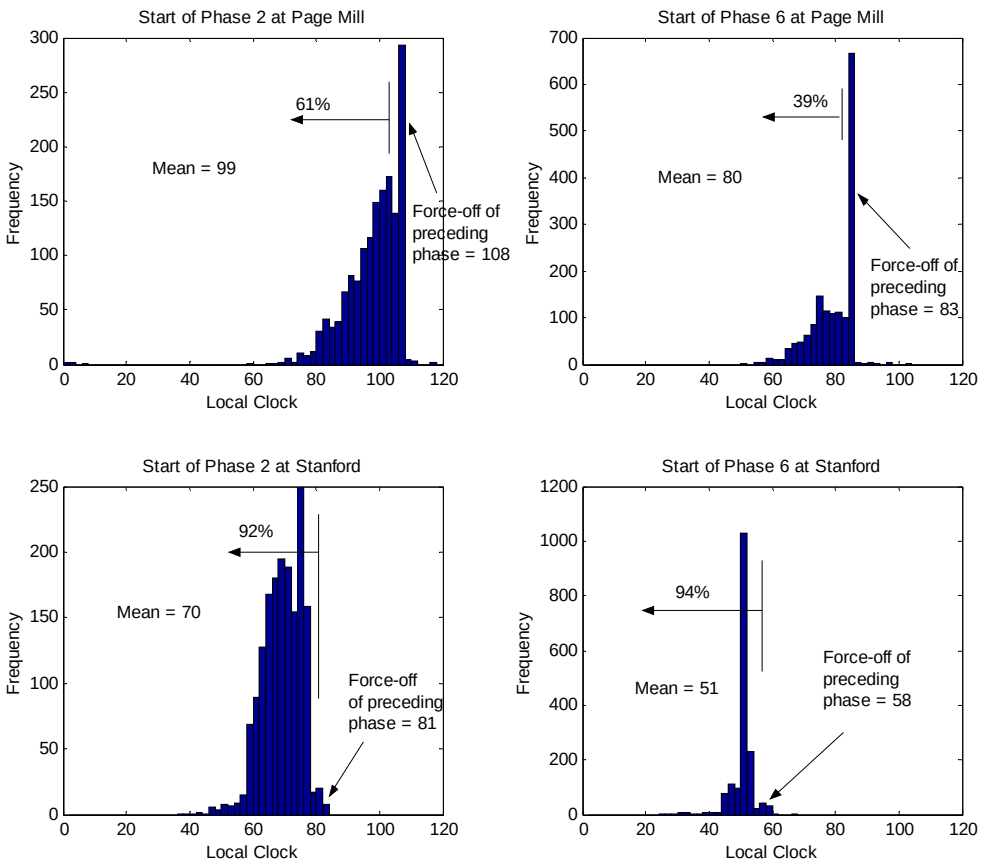


Figure 6-2. Probability of early-return-to-green at two intersections

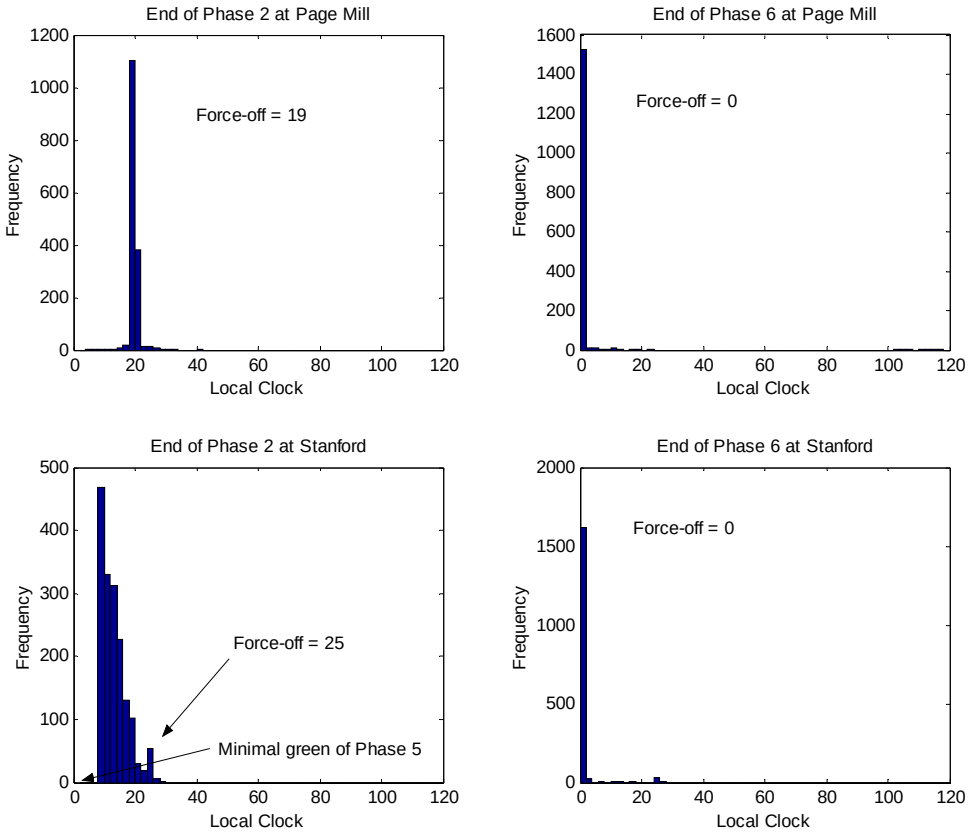


Figure 6-3. Uncertain termination of green at two intersections

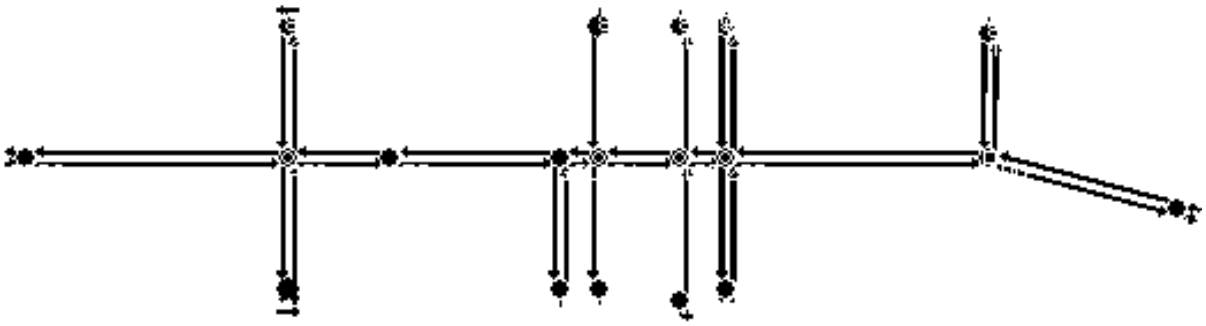


Figure 6-4. A snapshot of the “ActCtrl Example” in CORSIM

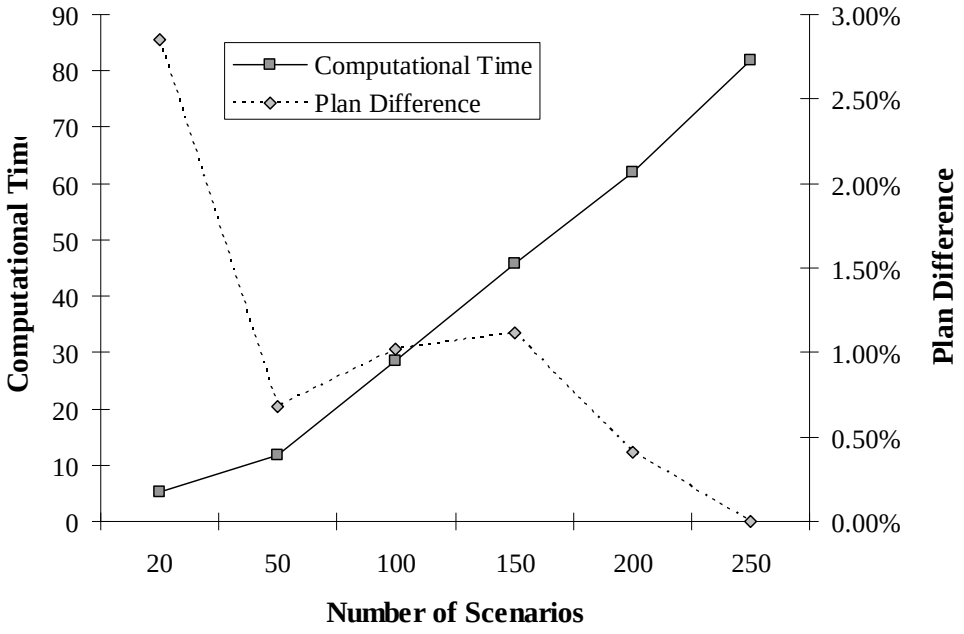


Figure 6-5. Computational time and plan difference

Table 6-1. Computation time and coordination plan difference

Number of scenario	Time*(sec)	Plan difference
10	2.88	169.3%
20	5.11	2.9%
50	11.67	0.7%
100	28.58	1.0%
150	45.67	1.1%
200	61.97	0.4%
250	81.81	0.0%

*: including the times for solving Little's formulation for each scenario

Table 6-2. Robust plan and nominal plan

	Offset (sec)					Cycle length (sec)
Intersection	2	3	4	5	6	
Robust plan	33	2	2	35	4	80
Nominal plan	53	74	72	47	68	77

Table 6-3. Monte Carlo simulation with normal distribution

Bandwidth (sec)	Mean	Worst case	90th percentile	90% CVaR	Change Mean	Worst case	90th percentile	90% CVaR
Nominal plan	28.7	17.0	23.8	40.1	-	-	-	-
Robust plan	35.8	21.8	30.6	34.5	24.7%	28.2%	28.6%	-14.0%

Table 6-4. Critical values for uniform distribution

Signal	Red time	
	Minimum duration (cycles)	Maximum duration (cycles)
# 1	0*	0.42
# 2	0	0.25
# 3	0	0.56
# 4	0	0.53
# 5	0	0.56
# 6	0	0.45

*: Minor phases skipped.

Table 6-5. Monte Carlo simulation with uniform distribution

Bandwidth (sec)	Mean	Worst case	90th percentile	90% CVaR	Change Mean	Worst case	90th percentile	90% CVaR
Nominal plan	36.6	17.8	24.3	54.1	-	-	-	-
Robust plan	44.3	22.6	31.0	46.9	21.0%	27.0%	27.6%	-13.3%

Table 6-6. Microscopic simulation results and hypothesis tests

Signal plan	Performance measure	Control delay (sec)	Travel time (sec)	Stop ratio
Robust plan	Mean	96.2	318.4	0.148
	Std. deviation	12.2	12.7	0.008
Nominal plan	Mean	122.6	346.8	0.176
	Std. deviation	9.9	12.4	0.010
Hypothesis test	*t* value	5.31	5.06	7.08
	F value	1.54	1.05	1.32

CHAPTER 7
CONCLUSIONS AND FUTURE WORK

7.1 Conclusions

This dissertation has presented a stochastic programming approach to deal with a variety of uncertainties associated with signal timing optimization for fixed-time and actuated traffic signals. Representing the uncertain parameter of interest as a number of scenarios and the corresponding probabilities of occurrence, the stochastic programming approach optimizes signal timings with respect to a set of high-consequence scenarios. The proposed approach is demonstrated in three applications.

The first application optimizes the timing of fixed-time signals along arterials under day-to-day demand variations or uncertain future traffic growth. By specifying scenarios as realizations of uncertain demand, a general signal optimization model is formulated to optimize the cycle length, green splits, offsets and phase sequences in an integrated manner. Considering a large number of binary variables in the formulation, a simulation-based GA is developed to solve the problem. It should be mentioned that the setting of the GA-based algorithm, such as the fitness function, may influence the quality of the final plan and the convergence speed. Numerical experiments are needed to fine-tune the setting. The simulation-based model is broadly applicable, particularly when the objective function is difficult or time-consuming to evaluate.

The second application expands the first one by further considering the impact of signal timing on vehicle emissions. A bi-objective signal optimization model is developed to make an explicit tradeoff between delays and the human emission exposure, considering the uncertainty of wind speed and direction. The modal emission approach is integrated with the cell

transmission model to compute the pollutant emission rates at each modeling step, followed by a cell-based Gaussian plume dispersion model to capture the pollutant dispersion to estimate the human emission exposure. A GA-based solution approach is employed to solve the large mixed-integer problem to come by a set of Pareto optimal signal timing plans, minimizing the total system delay and the mean excess exposure simultaneously. The solutions form an efficient frontier that represents explicit tradeoffs between these two objective functions. Based on the decision maker's consideration of social costs of delays and emissions, an optimal timing plan can be selected.

The last application synchronizes actuated signals along arterials (the offsets and cycle length) for smooth and stable progression under uncertain traffic conditions, addressing the issue of uncertain starts/ends of green of sync phases. By specifying scenarios as realizations of uncertain red times of sync phases, a robust counterpart of Little's formulation is formulated as another MILP. The model can be easily solved using the state-of-the-art solvers, and the computational time only increases polynomially as the number of scenarios increases. The approach can be used to either design a new coordination plan for implementation or fine-tune the plan offline after implementation.

These three applications validate the stochastic programming approach for optimizing signal timings under uncertainty. The resulting timing plans have been demonstrated to perform more robustly and effectively in an uncertain environment, thereby leading to more reliable and sustainable mobility.

7.2 Future Work

With the same stochastic programming framework, an immediate follow-up work would be to develop an integrated model to directly optimize the settings of actuated signals, including not only the cycle length and offsets as in Chapter 6, but also other general settings like

minimum and maximum greens, green extensions and phase sequences. Developing an exact or approximated mathematical representation of the actuated control logic can be a big challenge.

It is also noted that the proposed stochastic programming approach is quite flexible to consider other types of uncertainties, e.g., the saturation flow rates. Future work may expand the current models to simultaneously address a variety of uncertainties from both the demand side and supply side. For example, by specifying a number of scenarios as realizations of certain saturation flow rates and uncertain demands simultaneously, another scenario-based stochastic program can be formulated to minimize some selected performance measure to obtain a robust timing plan.

Another potential future work is to expand the proposed models for more sophisticated network configurations, such as grid networks. The major issue to be resolved will be the coordination among multiple signals in the complex networks. The embedded traffic flow model will need to be enhanced for this purpose.

The simulation-based genetic algorithm proposed in the dissertation is broadly applicable, particularly when the objective function or constraints are difficult to evaluate. However, as the problem size increases, the algorithm will easily become time consuming. Future work may explore how to further improve the efficiency of the algorithm through, e.g., fine-tuning the fitness function used in the algorithm.

Given the promising experiment results in these three applications, future work may also be a field operation test. A software package can be also developed to include these three models, with necessary user interfaces for parameter input and result output. To apply these models, users should provide the necessary data for both the supply side and the demand side. At this stage, users will also have to code the network into the model, and estimate the distributions of the

uncertain parameters of interest from the raw data. In the future, two modules can be developed to facilitate the usage of the models: one is to estimate the probability distributions of the parameters from the input raw data with multiple choices of distribution types, and the other is to automatically generate network representations. The latter is particularly useful if the cell transmission model is used as the underlying traffic flow model.

LIST OF REFERENCES

Abbas, M., Bullock, D., Head, L., 2001. Real-time offset transitioning algorithm for coordinating traffic signals. Transportation Research Record 1748, 26–39.

Agbolosu-Amison, S.J., Park, B., 2009. Performance evaluation of dynamic gap-out feature at actuated traffic signal controller using a simulation based optimization method. The 88th Annual Meeting of the Transportation Research Board, (Compendium of Papers DVD-Rom), No.09-2802.

Allsop, R.E., 1968. Selection of offsets to minimize delay to traffic in a network controlled by fixed-time signals. Transportation Science 2(1), 1–13.

Allsop, R.E., 1971. Delay-minimizing settings for fixed-time traffic signals at a single road junction. IMA Journal of Applied Mathematics 8(2), 164–185.

Allsop R. E., 1972. Sensitivity of delay at a fixed time traffic signal to small errors in the observations used for calculating signal settings. In Traffic Flow and Transportation (Edited by G. F. Newell), 253–268, Elsevier, London.

Barth, M., An, F., Younglove, T., Scora, G., Levine, C., Ross, M., Wenzel, T., 2000. Comprehensive Modal Emission Model (CMEM), Version 2.0 User's Guide, Riverside, CA.

Ben-Tal, A., Nemirovski, A., 2002. Robust optimization—methodology and applications. Mathematical Programming, Series B, 92, 456–480.

Bertsimas D., Sim, M., 2003. Robust discrete optimization and network flows. Mathematical Programming, Series B, 98, 49–71.

Boillot, F., Blosseville, J.M., Lesort, J. B., Motyka, V., Papageorgiou, M., Sellam, S., 1992. Optimal signal control of urban traffic networks. In Proceedings of the 6th IEEE International Conference on Road Traffic Monitoring and Control, 75–79.

Brilon, W., Geistefeldt, J., Regler, M., 2005. Reliability of freeway traffic flow: a stochastic concept of capacity. In Proceedings of the 16th Symposium on Transportation and Traffic Theory, 125–144.

Brooke, A., Kendirck, D., Meeraus, A., 1992. GAMS: A User's Guide. The Scientific Press, South San Francisco, California.

Cambridge Systematics, 2009. Moving Cooler: An Analysis of Transportation Strategies for Reducing Greenhouse Gas Emissions. Urban Land Institute, Washington, D.C.

Cernuschi, S., Giugliano, M., Cemin, A., Giovannini, I., 1995. Modal analysis of vehicle emission factors. The Science of the Total Environment, 169, 175–183.

Chang, E.C.P., 1996. Guidelines for actuated controllers in coordinated systems. Transportation Research Record, 1554, 61–73.

Chang, E., Lei, J.C., Messer, C.J., 1988. Arterial Signal Timing Optimization using PASSER-II-87 Microcomputer User's Guide. TTI Research Report 467-1, Texas A&M University, College Station, Texas.

Chang, T.H., Lin, J.T., 2000. Optimal signal timing for an oversaturated intersection. Transportation Research, Part B, 34, 471–491.

Chen, G., Daskin, M.S., Shen, Z.J.M., Uryasev, S., 2006. The α-reliable mean-excess regret model for stochastic facility location modeling. Naval Research Logistics, 53, 617–626.

Coelho, M.C., Farias, T.L., Rouphail, N.M., 2005. Impact of speed control traffic signals on pollutant emissions. Transportation Research, Part D, 10, 323–340.

Comert, G., Cetin, M., Nichols, A. P., 2009. Incorporating queue length measurements into actuated signal control: evaluation of efficiency benefits at an intersection. The 88th Annual Meeting of the Transportation Research Board, (Compendium of Papers DVD-Rom), No. 09-2566.

CPLEX, Version 9.0., 2004. CPLEX Optimization Inc., Nevada.

Daganzo, C.F., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research, Part B, 28, 269–287.

Daganzo, C.F., 1995. The cell transmission model, part II: network traffic. Transportation Research, Part B, 29, 79–93.

Delucchi, M.A., Murphy, J.J., McCubbin, D.R., 2002. The health and visibility cost of air pollution: a comparison of estimation methods. Journal of Environmental Management, 64, 139–152.

Federal Highway Administration, 1996. Traffic Control Systems Handbook. U.S. Department of Transportation. Publication Number: FHWA-SA-95-032.

Federal Highway Administration, 2008. Traffic Signal Timing Manual. U.S. Department of Transportation. Publication Number: FHWA-HOP-08-024

Federal Highway Administration, 2009. Signal Timing under Saturated Conditions. http://ops.fhwa.dot.gov/publications/fhwahop09008/chapter2.htm

Flyvbjerg, B., Holm, M.K.S., Buhl, S.L., 2005. How (in)accurate are demand forecasts in public works projects: the case of transportation. Journal of the American Planning Association, 71 (2), 131–146.

Flyvbjerg, B., Holm, M.K.S., Buhl, S.L., 2006. Inaccuracy in traffic forecasts. Transport Reviews, 26 (1), 1–24.

Frey, H.C., Rouphail, N.M., Unal, A., Colyar, J.D., 2001. Emission Reduction Through Better Traffic Management: An Empirical Evaluation Based Upon On-Road Measurements. Report No. FHWA/NC/2002-001, North Carolina Department of Transportation, USA.

Frey, H.C., Unal, A., Chen, J., 2002. Recommended Strategies for On-Board Emission Data Analysis and Collection for the New Generation Model, Prepared by NC State University for Office of Transportation and Air Quality, U.S. Environmental Protection Agency, Ann Arbor, Michigan.

Gartner, N.H., 2002. Development and implementation of an adaptive control strategy in a traffic signal network: the virtual-fixed-cycle approach. In Proceeding of 15th International Symposium on Transportation and Traffic Theory, 137–155.

Gartner, N.H., Assmann, S.F., Lasaga, F., Hou, D.L., 1991. A multi-band approach to arterial traffic signal optimization. Transportation Research, Part B, 25, 55–74.

Gartner, N.H., Stamatiadis, C., 2002. Arterial-based control of traffic flow in urban grid networks. Mathematical and Computer Modeling, 35, 657–671.

Gartner, N.H., Stamatiadis, C., 2004. Progression optimization featuring arterial-and route-based priority signal networks. ITS Journal, 8, 77–86.

Gazis, D.C., 1964. Optimum control of a system of oversaturated intersections. Operation Research, 12(6) 815–831.

Geistefeldt, J., 2008. Empirical relation between stochastic capacities and capacities obtained from the speed-flow diagram. In Proceedings of the Symposium on the Fundamental Diagram: 75 years, Woods Hole, MA.

Gettman, D., Shelby, S.G., Head, L., Bullock, D.M., Soyke, N., 2007. Data-driven algorithms for real-time adaptive tuning of offsets in coordinated traffic signal systems. Transportation Research Record, 2035, 1–9.

Head, L., Gettman, D., Wei, Z., 2006. Decision model for priority control of traffic signals. Transportation Research Record, 1978, 169–177.

Hensher, D., Button, K., Sperling D., 2001. Environment Protection. In Handbook of Transport Systems and Traffic Control. Institute of Transportation Studies, University of California, Davis, Research Report UCD-ITS-RR-01-19.

Heydecker, B., 1987. Uncertainty and variability in traffic signal calculations. Transportation Research, Part B, 21, 79–85.

Hunt, P.B., Robertson, D.I., Bretherton, R.D., Winton, R.I., 1981. SCOOT - a Traffic Responsive Method of Co-ordinating Signals. TRL Laboratory Report 1014.

Jongenotter, E., Monsma, S., 2002. Green logic and the benefits for traffic-actuated control – the Dutch experience. Traffic Engineering & Control, 43(9), 351–353.

Jovanis, P.P., Gregor, J.A., 1986. Coordination of actuated arterial traffic signal systems. Journal of Transportation Engineering, 112(4), 416–432.

Krokhmal, P., Palmquist, J., Uryasev, S., 2002. Portfolio optimization with conditional value-at-risk objective and constraints. Journal of Risk, 4, 46–68.

Lan C.J., 2004. New optimal cycle length formulation for pre-timed signals at isolated intersections. Journal of Transportation Engineering, 130(5), 637–647.

Li, H., Prevedouros, P.D., 2004. Traffic adaptive control for oversaturated isolated intersections: model development and simulation testing. Journal of Transportation Engineering, 130(5), 594–601.

Li, X., Li, G., Pang, S., Yang, X., Tian, J., 2004. Signal timing of intersections using integrated optimization of traffic quality, emissions and fuel consumption: a note. Transportation Research, Part D, 9, 401–407.

Lin, J., Ge, Y.E., 2006. Impacts of traffic heterogeneity on roadside air pollution concentration. Transportation Research, Part D, 11, 166–170.

Lin, W., Wang, C., 2004. An enhanced 0-1 mixed integer LP formulation for traffic signal control. IEEE Transactions on Intelligent Transportation Systems, 5, 238–245.

Little, J.D., 1966. The synchronization of traffic signals by mixed-integer linear programming. Operations Research, 14(4), 568–594.

Little, J.D., Kelson, M.D., Gartner, N.H., 1981. MAXBAND: A program for setting signals on arteries and triangular networks. Transportation Research Record, 795, 40–46.

Liu, H., Bhimireddy, S., 2009. Evaluation of integrated platoon-priority and advance warning flasher system at high-speed intersections. At The 88th Annual Meeting of the Transportation Research Board, (Compendium of Papers DVD-Rom) 09-2111

Lo, H., 1999. A novel traffic signal control formulation. Transportation Research, Part A, 44, 436–448.

Lo, H., 2001. A cell-based traffic control formulation: strategies and benefits of dynamic timing plan. Transportation Science, 35, 148–164.

Lo, H., Chang, E., Chan. Y. C., 2001. Dynamic network traffic control. Transportation Research, Part A, 35, 721–744.

Lou, Y., Yin, Y., Lawphongpanich, S., 2009. Robust congestion pricing with boundedly rational user equilibrium. Transportation Research, Part B, 44(1), 15–28.

Lowrie, P., 1982. The sydney coordinate adaptive traffic system – principles, methodology, algorithms. In Proceedings of the International Conference on Road Traffic Signaling. Institution of Electrical Engineers, London, 67–70.

Luyanda, F., Gettman, D., Head, L., Shelby, S., Bullock, D., Mirchandani, P., 2003. ACS-lite algorithmic architecture: applying adaptive control system technology to closed-loop traffic signal control systems. design guidelines for deploying closed loop systems. Transportation Research Record 1856, 175–184.

Martinez, R., Wu, W., McNeill, K., Deal, J., Haynes, T., Bradford, D., 2003. Hardware and software-in-the-loop techniques using the OPNET modeling tool for JTRS developmental testing. IEEE Military Communications Conference, 1, 469–474.

Matson, T.M., Smith, W.S., Hurd, F.W., 1955. Traffic Engineering. MacGraw-Hill, New York.

Messer, C.J., Hogg, G.L., Chaudhary, N.A., Chang, E., 1987. Optimization of Left Turn Phase Sequence in Signalized Networks using MAXBAND 86, Vol. 1 Summary Report. Report Number: FHWA/RD-87/109, U.S. Department of transportation.

Michalopoulos, P.G., Stephanopoulos G., 1977. Oversaturated signal systems with queue length constraints—I : single intersection. Transportation Research, 11(6), 413–421.

Miller, A.J., 1965. A computer control system for traffic networks. In Proceedings of the 2nd International Symposium on the Theory of Road Traffic Flow, Paris, OECD.

Minderhoud, M.M., Botma H., Bovy P.H.L., 1997. Assessment of roadway capacity estimation methods. Transportation Research Record 1572, 59–67.

Mirchandani, P. B., Zou, N., 2007. Queuing models for analysis of traffic adaptive signal control. IEEE Transactions on Intelligent Transportation System, 8(1), 50–59.

Morgan J.T., Little, J.D., 1964. Synchronizing traffic signals for maximal bandwidth. Operation Research, 12(6), 896–912.

Newell, G.F., 1998. The rolling horizon scheme of traffic signal control. Transportation Research, Part A, 32(1), 39–44.

NTOC, 2007. 2007 National Traffic Signal Report Card Technical Report.

Oak Ridge National Laboratory, 2004. Temporary Losses of Highway Capacity and Impacts on Performance: Phase 2. Report Number: ORNL/TM-2004/209, Oak Ridge, TN, US.

Osyczka, A., Kundu, S., 1996. A modified distance method for multi-criteria optimization, using genetic algorithms. Computers and Industrial Engineering, 30(4), 871–882.

Park, B., Messer, C.J., Urbanik, T., 2000. Enhanced genetic algorithm for signal-timing optimization of oversaturated intersections. Transportation Research Record 1727, 32–41.

Park, B., Yun, I., Ahn, K., 2009. Stochastic optimization of sustainable traffic signal control. International Journal of Sustainable Transportation, 3, 263–284.

Pavlis, Y., Recker, W.A., 2009. Mathematical logic approach for the transformation of the linear conditional piecewise functions of dispersion-and-store and cell transmission traffic flow models into mixed-integer form. Transportation Science, 43, 98–116.

Ponzlet, M., 1996. Dynamik der leistungsfähigkeiten von autobahnen (dynamics of freeway capacity). Schriftenreihe des Lehrstuhls fuer Verkehrswesen der Ruhr-Universitaet Bochum, 16. Bochum.

Poole, R., 2009. Reducing greenhouse gas emissions without reducing mobility. Reason Foundation. http://reason.org/news/show/1007715.html

Rakha, H., Ahn, K., Trani, A., 2004. A development of VT-micro model for estimating hot stabilized light duty vehicle and truck emissions. Transportation Research, Part D, 2004, 9, 49–74.

Rakha, H., Van Aerde, M., Ahn, K., Trani, A., 2000. Requirements for evaluating traffic signal control impacts on energy and emissions based on instantaneous speed and acceleration measurements. Transportation Research Record, 1738, 56–67.

Reiss, R.D., Thomas, M., 2007. Statistical Analysis of Extreme Values, 3rd edition. Birkhauser.

Ribeiro, P.C.M., 1994. Handling traffic fluctuation with fixed-time plans calculated by TRANSYT. Traffic Engineering and Control, 35, 365–366.

Robertson, D.I., 1969. TRANSYT -- a Traffic Network Study Tool. Report Number: TRRL-LR-253, Transport and Road Research Laboratory.

Rockafellar, R.T., 2001. Optimization under Uncertainty: Lecture Notes. University of Washington

Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. Journal of Risk, 2, 21–41.

Rockafellar, R.T., Uryasev, S., 2002. Conditional value-at-risk for general loss distribution. Journal of Banking and Finance, 26, 1446–1471.

Rockafellar, R.T., Wets, R.J.B., 1991. Scenarios and policy aggregation in optimization under uncertainty. Mathematics of Operations Research, 16(1), 119–147.

Sahinidis, N.V., 2004. Optimization under uncertainty: state of the art and opportunities. Computation and Chemical Engineering, 28, 971–983.

Schrank, D., Lomax, T., 2009. Urban Mobility Report. Texas Transportation Institute, Texas A&M University.

Shoup, G.E., Bullock, D., 1999. Dynamic offset tuning procedure using travel time data. Transportation Research Record 1683, 84–94.

Skabardonis, A., 1996. Determination of timings in signal systems with traffic-actuated controllers. Transportation Research Record 1554, 18–26.

Smith, M., Clegg, J., Yarrow R., 2001. Modeling traffic signal control. In Handbook of Transport Systems and Traffic Control (edited by: Button K.J., Hensher, D.A.) Elsevier.

Smith, B.L., Scherer, W.T., Hauser, T.A., Park., B.B., 2002. Data-driven methodology for signal timing plan development: a computational approach. Computer-Aided Civil and Infrastructure Engineering, 17, 387–395.

Transportation Research Board, 2000. Highway Capacity Manual. National Research Council, Washington, D.C.

Turner, D.B., 1994. Workbook of Atmospheric Dispersion Estimates: an Introduction to Dispersion Modeling, CRC Press.

USDA Natural Resources Conservation Service, NRCS National Water & Climate Center: Climate Products: Climate Data: Wind Rose Data. ftp://ftp.wcc.nrcs.usda.gov/downloads/climate/windrose/california/san_francisco/

Vincent, R.A., Peirce, J.R., 1988. 'MOVA': Traffic Responsive, Self-Optimising Signal Control for Isolated Intersections. Report number 170, Transport and Road Research Laboratory.

Viti, F., Van Zuylen, H.J., 2009. A probabilistic model for traffic at actuated control signals. Transportation Research, Part C, in Press.

Wallace, C.E., Courage, K.G., Hadi, M.A., Gan, A.G., 1998. TRANSYT-7F User's Guide. University of Florida, Gainesville, FL.

Webster, F.V., 1958. Traffic Signal Settings. Road Research Technical Paper, 39, Her Majesty's Stationary Office, London, U.K..

Yin, Y., 2002. Multi-objective bi-level optimization for transportation planning and management problems. Journal of Advanced Transportation, 36(1), 93–105.

Yin, Y., 2008a. A scenario-based model for fleet allocation of freeway service patrols. Network and Spatial Economics, 8(4), 407–417.

Yin, Y., 2008b. Robust optimal traffic signal timing. Transportation Research, Part B, 42, 911–924

Yin, Y., Lawphongpanich, S., 2006. Internalizing emission externality on road networks. Transportation Research, Part D, 11, 292–301.

Yin, Y., Li, M., Skabardonis, A., 2007. An offline offset refiner for coordinated actuated signal control system. Journal of Transportation Engineering, 133(7), 426–432.

Yu, L., Jia, S.C., Shi, Q.Y., 2009. Research on transportation-related emissions: current status and future directions. Journal of The Air and Waste Management Association, 59, 183–195.

BIOGRAPHICAL SKETCH

Lihui Zhang was born in Zhoushan, China, in 1984. He studied structure engineering in Tsinghua University and received his bachelor's degree in 2006. Then, he joined the PhD program in transportation engineering in the University of Florida, under the supervision of Dr. Yafeng Yin. During his PhD study, Lihui Zhang worked on projects related to travel demand modeling, signal timing optimization and transportation network modeling. He was also the teaching assistant for two major transportation courses.

Lihui Zhang has co-authored four papers in leading transportation journals and conference proceedings, and has delivered six presentations in various conferences. His main research interests include transportation system modeling and optimization, traffic control and operation, traffic simulation, and travel demand modeling.