

Final Report
to the
CENTER FOR MULTIMODAL SOLUTIONS FOR CONGESTION MITIGATION
(CMS)

CMS Project Number: 2008-003

CMS Project Title: Simulation-Based Robust Optimization for
Signal Timing and Setting

for period 3/1/2008 to 8/31/2009

from Lihui Zhang, Yafeng Yin and Scott Washburn
Department of Civil and Coastal Engineering
University of Florida
Email: yafeng@ce.ufl.edu (Yafeng Yin)

Date prepared 12/30/09



DISCLAIMER AND ACKNOWLEDGMENT

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This document is disseminated under the sponsorship of the Department of Transportation University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

This work was sponsored by a grant from the Center for Multimodal Solutions for Congestion Mitigation, a U.S. DOT Tier-1 grant-funded University Transportation Center. The authors wish to thank Meng Li at California PATH, University of California at Berkeley for providing the data used in one of the numerical examples. The report also benefits from the discussions with Yingyan Lou at University of Alabama and Farid AitSahlia at University of Florida.



TABLE OF CONTENTS

DISCLAIMER AND ACKNOWLEDGMENT	i
LIST OF TABLES	iii
LIST OF FIGURES	iv
ABSTRACT	v
EXECUTIVE SUMMARY	vi
CHAPTER 1 BACKGROUND	1
CHAPTER 2 CELL TRANSMISSION MODEL	6
Model Introduction	6
Encapsulating CTM in Signal Timing Optimization	8
CHAPTER 3 ENHANCED DETERMINISTIC SIGNAL OPTIMIZATION MODEL	9
Objective Function	9
Constraints	11
CHAPTER 4 STOCHASTIC SIGNAL OPTIMIZATION MODEL	19
CHAPTER 5 NUMERICAL EXAMPLES	22
Simulation-Based Genetic Algorithm	22
Numerical Example I	27
Numerical Example II	33
CHAPTER 6 CONCLUDING REMARKS	40
REFERENCE	41



LIST OF TABLES

<u>Table</u>	<u>Page</u>
3-1 CPF Result under Different 0-1 Combinations.....	11
5-1 Traffic Data for Three-Node Network.....	29
5-2 Signal Plans for Three-Node Network.....	31
5-3 Two-Way Bandwidth Results for Three-Node Netowrk.....	32
5-4 CORSIM Result for Three-Node Network.....	33
5-5 Traffic Data for El Camino Real Arterial	35
5-6 Signal Plans for El Camino Real Arterial	37
5-7 CORSIM Result for El Camino Real Arterial	38



LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1-1 Day-to-day AM-peak hourly flow rate at one intersection in Gainesville, FL.....	2
2-1 Piecewise linear $q - k$ relationship	7
3-1 Cell configurations.....	10
3-2 Interpretation of the objective function.....	10
3-3 Traffic intersection configurations.....	15
3-4 NEMA phasing structure for a four-way intersection.....	16
3-5 NEMA phasing structure for a t -intersection.....	17
4-1 Illustration of concept of mean excess delay	21
5-1 Flow-chart of simulation-based GA.....	23
5-2 Configuration of the chromosome	25
5-3 Cell representation of the three-node network.....	28
5-4 Convergence of GA under both traffic conditions.....	30
5-5 Time-space diagram of the three-node network	32
5-6 A snapshot of the three-node network in CORSIM.....	33
5-7 Cell representation of El Camino Real Arterial	34
5-8 Convergence of GA with both fitness functions.....	36
5-9 A snapshot of the EL Camino Real Arterial.....	39



ABSTRACT

The performance of signal timing plans obtained from traditional approaches for pre-timed (fixed-time or actuated) control systems is often unstable under fluctuating traffic conditions. This report develops a general approach for optimizing the timing of pre-timed 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, particularly when the optimization horizon is long and the network size is large. Consequently, a simulation-based genetic algorithm is developed to solve the model. The model and algorithm are validated and verified in two networks. It is demonstrated that the resulting robust timing plans perform better against high-consequence scenarios without losing optimality in the average sense. More specifically, the plans reduce substantially the mean excess delay across the high-consequence scenarios without compromising the average delay across all scenarios under both congested and uncongested traffic conditions.



EXECUTIVE SUMMARY

This report presents a methodology to determine robust signal timings against day-to-day demand fluctuations or uncertain future traffic growth. Compared with those from conventional timing approaches, robust timing plans are expected to perform better under high-demand scenarios without compromising the average performance across all possible demand scenarios. Although the methodology is applicable more widely, this report is focused on developing robust timing models for pre-timed (fixed-time or actuated) signals along arterials.

Traditional timing approaches do not proactively consider traffic demand uncertainty, and thus the resulting control performance is often unstable under fluctuating traffic conditions. (Although actuated signals can respond to traffic fluctuations to a certain degree, the underlying timing plan still plays an important role in determining the efficiency of the control, especially over a coordinated semi-actuated signal-controlled arterial.) The underlying premise of the ubiquitous time-of-day control is that traffic patterns remain essentially the same throughout a time-of-day period. However, in real life, traffic arrivals to intersections can vary significantly even for the same time of day and day of week. If the traffic arrivals vary significantly, optimizing signal timing with respect to the average may cause a considerable amount of additional delay. Even if the arrivals do not vary much, use of the average flow may lead to a substantial loss in the stability of control performance, thereby making motorists' travel times unpredictable and unreliable.

Practically motorists and traffic engineers may be more concerned with worst-case scenarios where substantial delay may occur. To address such a risk-averse attitude on one hand and avoid being too conservative on the other hand, this report optimizes signal timings against a set of worst-case or high-consequence scenarios. More specifically, given a set of demand scenarios and their corresponding probability of occurrence, this report formulates a stochastic programming model to simultaneously determine cycle length, green splits, phase sequences and offsets to minimize the mean of the delays exceeding the α -percentile (e.g., 90th percentile) of the entire delay distribution, i.e., mean excess delay. 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, particularly when the optimization horizon is long and the network size is large. This report develops a simulation-based genetic algorithm to solve the model. The model and algorithm are tested on two networks and the resulting robust timings are compared with traditional timing plans via a CORSIM simulation study. It is demonstrated that the robust timing plans substantially reduce the mean excess delay across the high-consequence scenarios without compromising the average delay across all scenarios under both congested and uncongested traffic conditions.



1. BACKGROUND

This report presents a methodology for designing robust signal timing plans for pre-timed (fixed-time or traffic-actuated) signals along urban arterials under demand uncertainty. A signal timing plan is called robust if its performance is less sensitive to fluctuations of traffic demands or it performs better against worst-case or high-consequence demand scenarios without compromising optimality in the average sense.

Traffic congestion is one of the most severe problems that threaten the economic prosperity and quality of life in many societies. According to a report by Federal Highway Administration, traffic congestion in the U.S. costs approximately \$200 billion a year in wasted gas and time, five percent of which is incurred by poor signal timing. Additionally, reduced idleness at intersections is likely to lead to significantly reduced greenhouse gas emissions. While recent research has primarily focused on developing real-time adaptive signal control systems, implementation of such systems on a large scale may be many years away, due to the associated high implementation and maintenance costs (1). Because a large segment of signal control systems in use today are still pre-timed, further improvements in the efficiency (e.g., delay per vehicle) and robustness (e.g., variance of delay per vehicle) of signal control systems can yield significant improvements in the management of traffic flows and mitigation of congestion.

The performance of signal timings obtained from traditional approaches for pre-timed control systems is often unstable under fluctuating traffic conditions. Pre-timed control systems in current practice 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 (2) or using optimization tools such as TRANSYT-7F (3). 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 to intersections can vary significantly even for the same time of day and day of week. As an example, Figure 1-1 displays hourly arrivals at two crossing streets, 34th Street and University Avenue, in Gainesville, Florida, during an a.m. peak on weekdays over a period of four months. The flows present significant day-to-day variations.

A consequent issue that traffic engineers may be confronted with is to determine the flows to use to optimize signal timings. This issue was hardly a concern in old days since the



data collection used to be resource demanding, and traffic data were only collected for a couple of days. As the advancement of portable-sensor and telecommunications technologies make high-resolution traffic data more readily available, chances for traffic engineers to raise such a question become more prevalent. This is particularly true in re-timing efforts for those closed-loop control systems with fiber optic connections. For example, in California, second-by-second returns of loop data are archived and can be obtained as a result of AB3418 (The California legislature passed legislation, Assembly Bill 3418, requiring all signal controllers purchased in the state after January 1, 1996, to be compliant with a standardized communication protocol).

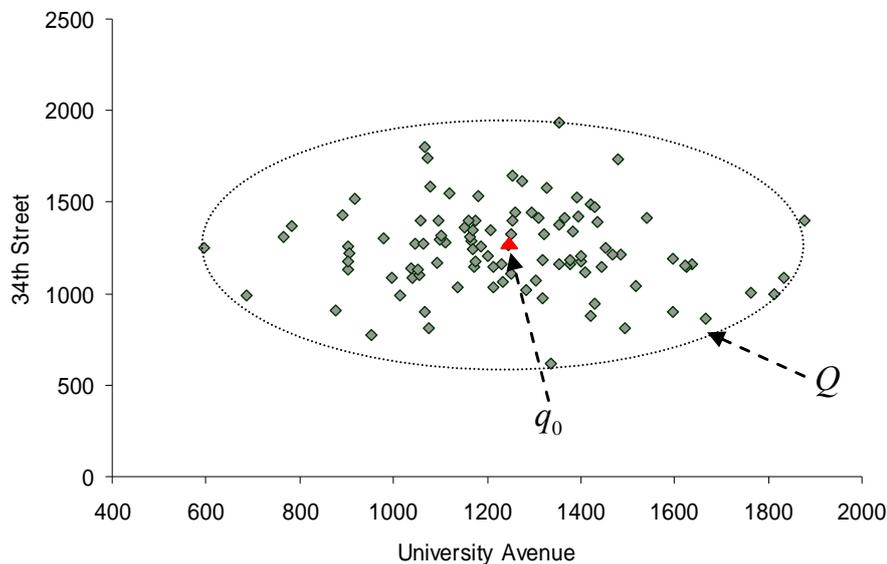


Figure 1-1. Day-to-day AM-peak hourly flow rate (in vehicle per hour) at one intersection in Gainesville, Fla.

Use of the average flows (i.e., q^0 in figure 1-1) may not be a sensible choice. Heydecker (4) 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 we observed in our preliminary investigation (5), it may still cause considerable losses in the performance against the worst-case scenarios or the stability of performance (robustness), thereby causing motorists' travel times 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. (1) suggested using 90th 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 (6).

There are two approaches to address this issue. One approach is to gradually adjust or refine signal settings in operations using the loop and signal status data to make signal operations more responsive to the traffic. The other approach is to explicitly consider the uncertainties in signal timing optimization prior to operations and develop robust timing plans that are able to tolerate those fluctuations and perform stably. The ACS-Lite program initiated by FHWA in 2002 (7, 8) adopts the first approach while this report considers the latter. Our preliminary investigation has demonstrated that, when compared with traditional approaches, robust timing plans may reduce standard deviations of delay per vehicle by 11.3 percent to 16.7 percent, and 28.2 percent to 44.6 percent under over-saturated and under-saturated conditions respectively, without worsening off the average performance at an isolated intersection (5). Those robust timing plans also allow slower deterioration of performance. We note that the signal timing process is normally time-consuming. Thus it is rarely repeated unless changes in traffic conditions are so significant that the system begins performing poorly. It has been estimated that traffic experiences an additional 3 percent to 5 percent delay per year as a consequence of not retiming signals as conditions evolve over time (7). Therefore, it is desirable to have timing plans that accommodate or tolerate these changes in traffic to a greater extent.

Since the seminal work of Webster (2), significant efforts have been devoted to improving signal timing for saturated isolated intersections, coordinated arterials and grid networks, etc. For example, Gazis and Potts (9) and Gazis (10) proposed semi-graphical methods to optimize the settings of an isolated signal and the system of two signals in oversaturated traffic conditions. The optimal signal setting involves values of the control variables that lie along edges of the control region, defined by the permissible ranges of the green phase splits. The authors also developed an analytical formulation of the method based on Pontryagin's theory. Robertson and Bretherton (11) described the evolution of an adaptive traffic control system SCOOT from the TRANSYT method, via introducing the improvement of the methodology for coordinating fixed-time signals. The advantage of SCOOT is that the system can measure the cycle flow profile in real time and then update the coordination plan in an on-online manner. The real-time capability makes dynamic traffic management possible. Gartner (12) proposed another 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. The phase splits of the signals in the coordination system are constrained by a virtual cycle length and virtual offsets updated based on real-time data.



Only a few studies have been conducted in the literature to directly address signal timing under flow fluctuations for pre-timed control systems. Heydecker (4) 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. Sensitivity analysis was carried out to test the benefit of taking into account the variability of arrival rate when optimizing the signal settings. Following the same notion, Ribeiro (13) proposed a novel technique called Grouped Network for using TRANSYT to calculate timing plans that are efficient even when demand is variable. Both studies focus on optimizing the average performance.

Motivated by recent developments in robust optimization (e.g., 14, 15), Yin (5) and Zhang and Yin (16) proposed a robust optimization approach for timing signals under uncertainty. Considering that travelers and traffic engineers may be more concerned with the adverse system performance and are less likely to complain if the delay is less than expected, the robust timing plans are designed to perform better against high-consequence scenarios. Yin (5) 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., 90th 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 percent and 11.3 percent respectively as well as the standard deviation of delay per vehicle by 12.0 percent and 16.3 percent respectively, without adversely affecting the average performance at a real-world intersection.

Zhang and Yin (16) proposed a robust model to synchronize actuated signals along arterials or at grid networks for smooth and stable progression under uncertain traffic conditions, mainly addressing the issue of uncertain (not fixed) starts/ends of the green of sync phases. The paper bases the model development on Little's mixed-integer linear programming (MILP) formulation (17), which maximizes 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, they defined 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. The numerical example shows that the resulting robust coordination plan is able to increase the worst-case and 90th percentile bandwidths by approximately 20 percent without affecting the average bandwidth.



One may argue that flow fluctuations have been considered by adding additional terms in delay formulas, such as those in 2000 Highway Capacity Manual, 1995 Canadian Capacity Guide, and 1981 Australian Capacity Guide etc. In other words, a signal timing that minimizes delay according to those delay formulas should be able to provide enough buffers to accommodate stochastic arrivals. We note that this report is concerned with day-to-day flow variations while the delay formulas were derived by considering intra-day random arrivals. For example, the second term of the Webster delay formula is obtained by assuming Poisson arrivals whose ratio of variance to mean is one. This assumption can be easily and significantly violated in day-to-day demand fluctuations. For instance, the ratio at the site of Figure 1-1 is as high as 47.7. We further note that traffic conditions under high-consequence demand scenarios are most likely over-saturated. Most of the existing timing approaches address under-saturated conditions and no coherent approach exists for oversaturated conditions, as concluded by NCHRP 3-38 (18). Consequently, two NCHRP research projects are investigating signal-timing strategies for over-saturated situations. This report also partly contributes to this ongoing research effort.



2. CELL-TRANSMISSION MODEL

2.1 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 (19, 20) in order to fully capture traffic dynamics, such as shockwaves, 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:

$$\frac{\partial k}{\partial t} + \frac{\partial q}{\partial x} = 0 \quad (1)$$

$$q = Q(k, x, t) \quad (2)$$

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

For a homogeneous roadway, Daganzo (19, 20) 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 2-1 shows the flow-density relationship in a piecewise linear diagram.

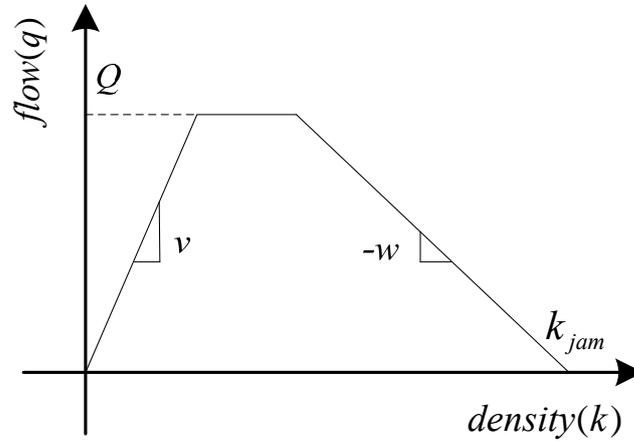


Figure 2-1. Piecewise linear $q-k$ relationship.

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:

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

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

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



2.2 ENCAPSULATING CTM IN SIGNAL TIMING OPTIMIZATION

Lin and Wang (21), Lo (22, 23) and Lo et al. (24) have successfully incorporated CTM in their signal timing optimization formulations.

Lin and Wang (21) 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) 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 (22, 23) and Lo et al. (24) developed dynamic signal control formulations based on CTM. By introducing binary variables, Equation (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 report 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.



3. 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 (25) 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 3-1 from (a) to (e). Each group has a different configuration to be discussed below.

3.1 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 3-2) between the cumulative arrival curves of the origin cells and the cumulative departure curves of the destination cells, expressed as the following linear function:

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

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.

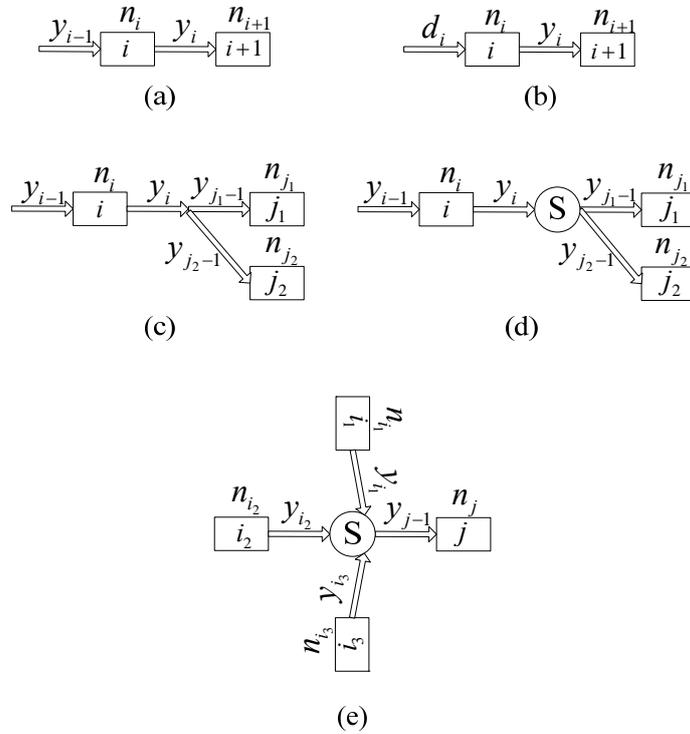


Figure 3-1. Cell configurations.

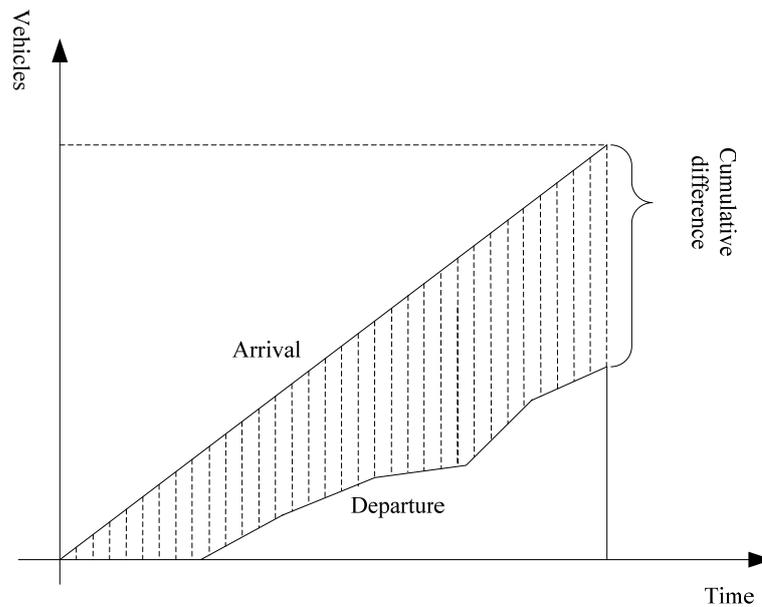


Figure 3-2. Interpretation of the objective function with cumulative vehicle curves.



3.2 CONSTRAINTS

3.2.1 Constraints for Ordinary Cells

The ordinary cells are those with only one inflow and one outflow as cell i in Figure 3-1(a). 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 (25) 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:

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

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

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

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

To see the equivalence, Table 3-1 enumerates all possible 0-1 combinations and evaluates the value of $y_i(t)$.

Table 3-1. CPF Result under Different 0-1 Combinations

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



3.2.2 Constraints for Origin Cells

The origin cells, as shown in Figure 3-1(b), 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:

$$\begin{aligned}
 n_i(t+1) &= n_i(t) + d_i(t) - y_i(t) \\
 (\chi_3 + \chi_4)U^- &\leq y_i(t) - n_i(t) \leq 0 \\
 (1 + \chi_3 - \chi_4)U^- &\leq y_i(t) - Q_{i,\max}(t) \leq 0 \\
 (1 - \chi_3 + \chi_4)U^- &\leq y_i(t) - Q_{i+1,\max}(t) \leq 0 \\
 (2 - \chi_3 - \chi_4)U^- &\leq y_i(t) - \omega \cdot [N_{i+1,\max} - n_{i+1}(t)] \leq 0
 \end{aligned}$$

3.2.3 Constraints for Destination Cells

The destination cells 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:

$$\begin{aligned}
 n_i(t+1) &= n_i(t) + y_{i-1}(t) - y_i(t) \\
 y_i(t) &= n_i(t)
 \end{aligned}$$

3.2.4 Constraints for Non-Signalized Diverge Cells

Non-signalized diverge occurs at certain roadway segments where the geometry or capacity changes and traffic diverge to different lanes for their respective destinations. Figure 3-1(c) 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:

$$\begin{aligned}
 y_{j_1} &= \beta_{j_1} \cdot y_i \\
 y_{j_2} &= \beta_{j_2} \cdot y_i \\
 \beta_{j_1} + \beta_{j_2} &= 1 \\
 (\chi_5 + \chi_6 + \chi_7)U^- &\leq y_i(t) - n_i(t) \leq 0 \\
 (1 + \chi_5 + \chi_6 - \chi_7)U^- &\leq y_i(t) - Q_{i,\max}(t) \leq 0 \\
 (1 + \chi_5 - \chi_6 + \chi_7)U^- &\leq y_i(t) - Q_{j_1,\max}(t) / \beta_{j_1} \leq 0
 \end{aligned}$$



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

3.2.5 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 3-1(d) sketches a configuration of the signalized diverge, where the sign S indicates a traffic signal. The constraints are as follows:

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

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 in Section 3.2.7.

3.2.6 Constraints for Signalized Merge Cells

Figure 3-1(e) 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:

$$(\chi_{11} + \chi_{12})U^- \leq y_{i_1}(t) - n_{i_1}(t) \leq 0$$

$$(1 + \chi_{11} - \chi_{12})U^- \leq y_{i_1}(t) - Q_{i_1}(t) \leq 0$$

$$(1 - \chi_{11} + \chi_{12})U^- \leq y_{i_1}(t) - Q_{j,\max}(t) \leq 0$$

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

Approach 2:

$$(\chi_{13} + \chi_{14})U^- \leq y_{i_2}(t) - n_{i_2}(t) \leq 0$$

$$(1 + \chi_{13} - \chi_{14})U^- \leq y_{i_2}(t) - Q_{i_2}(t) \leq 0$$

$$(1 - \chi_{13} + \chi_{14})U^- \leq y_{i_2}(t) - Q_{j,\max}(t) \leq 0$$

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

Approach 3:

$$(\chi_{15} + \chi_{16})U^- \leq y_{i_3}(t) - n_{i_3}(t) \leq 0$$

$$(1 + \chi_{15} - \chi_{16})U^- \leq y_{i_3}(t) - Q_{i_3}(t) \leq 0$$

$$(1 - \chi_{15} + \chi_{16})U^- \leq y_{i_3}(t) - Q_{j,\max}(t) \leq 0$$

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

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

3.2.7 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:

If $b(k, p, c) < t \leq e(k, p, c)$, then $Q_i(t) = s$; otherwise, $Q_i(t) = 0$,

where s is saturation flow rate;

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

$e(k, p, c)$ 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:

$$-U \cdot z_1(p, t) + \varepsilon \leq t - e(p) \leq U \cdot [1 - z_1(p, t)]$$

$$-U \cdot z_2(p, t) \leq b(p) - t \leq U \cdot [1 - z_2(p, t)] - \varepsilon$$



$$z_1(p,t) + z_2(p,t) - z(p,t) = 1$$

$$Q_i(t) = (z_1(p,t) + z_2(p,t) - 1) \cdot s$$

$$\sum_p (z_1(p,t) + z_2(p,t)) \leq 2$$

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.

3.2.8 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 3-3: (a) Four-way intersection and (b) T-intersection.

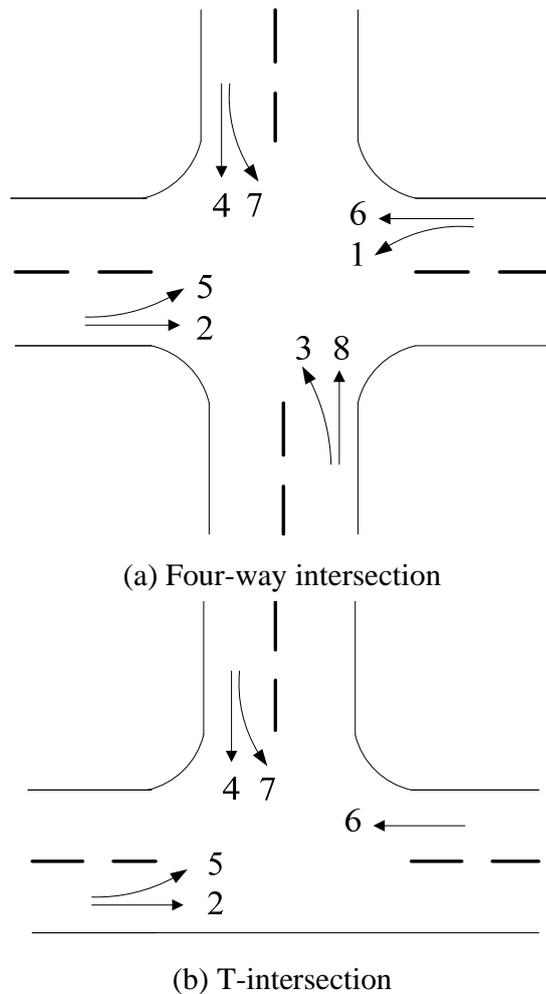
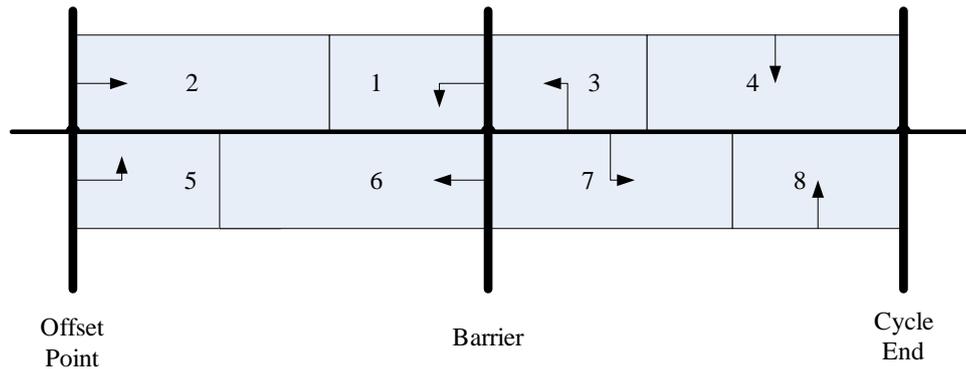


Figure 3-3. Traffic intersection configurations.

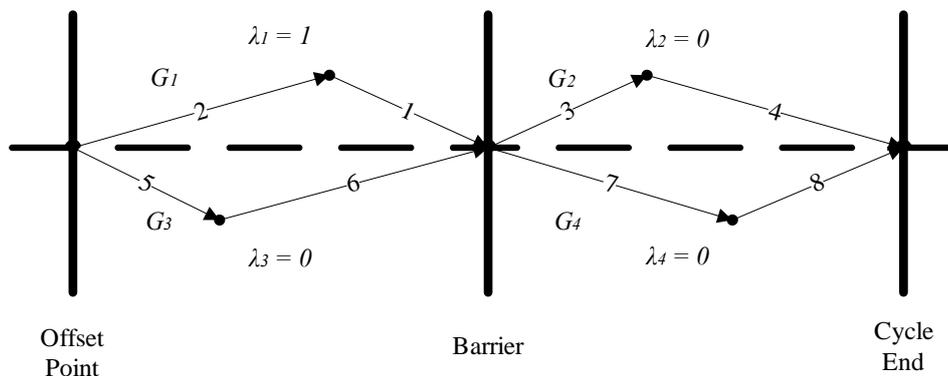


3.2.8.1 Four-Way Intersection

Figure 3-4(a) 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) NEMA phasing



(b) Transformation of NEMA phasing

Figure 3-4. NEMA phasing structure for a four-way intersection.

A binary variable is introduced for each portion as shown in Figure 3-4(b). 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:

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



$$\begin{aligned}
 e(k, "1", c) &= b(k, "1", c) + g(k, "1") \\
 b(k, "2", c) &= (1 - \lambda_1) \cdot o(k) + (1 - \lambda_1) \cdot l \cdot (c - 1) + \lambda_1 \cdot e(k, "1", c) \\
 e(k, "2", c) &= b(k, "2", c) + g(k, "2") \\
 g(k, "1") + g(k, "2") &= h(k)
 \end{aligned}$$

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, and 7 and 8. Once the value of $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is determined, the left-turn leading and lagging information can be obtained explicitly. Figure 3-4(b) presents a particular phase sequence corresponding to $\lambda_1 = 1$, $\lambda_2 = 0$, $\lambda_3 = 0$ and $\lambda_4 = 0$.

3.2.8.2 T-Intersection

T-intersection can be modeled the same way as the four-way intersection but is much simpler.

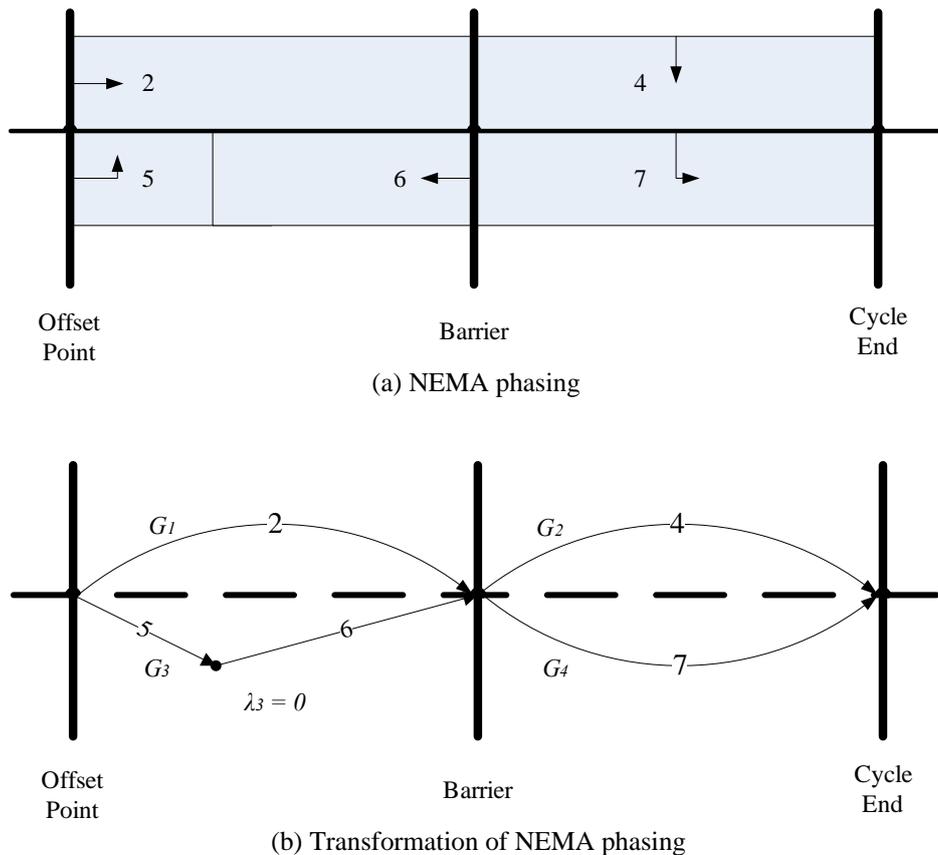


Figure 3-5. NEMA phasing structure for a T-intersection.



Figure 3-5 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:

$$\begin{aligned}
 b(k, "5", c) &= \lambda_3 \cdot o(k) + \lambda_3 \cdot l \cdot (c - 1) + (1 - \lambda_3) \cdot e(k, "6", c) \\
 e(k, "5", c) &= b(k, "5", c) + g(k, "5") \\
 b(k, "6", c) &= (1 - \lambda_3) \cdot o(k) + (1 - \lambda_3) \cdot l \cdot (c - 1) + \lambda_3 \cdot e(k, "5", c) \\
 e(k, "6", c) &= b(k, "6", c) + g(k, "6") \\
 g(k, "5") + g(k, "6") &= h(k)
 \end{aligned}$$

3.2.9 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 $(l^*, o^*, \lambda^*, g^*)^T$, where l^* , o^* , λ^* and g^* are vectors of optimal cycle length, offsets, phase sequences and green splits.



4. STOCHASTIC SIGNAL OPTIMIZATION MODEL

In the above deterministic case, it is assumed that traffic demand is fixed within the optimization horizon. In this chapter, a stochastic programming model is presented to accommodate day-to-day demand variations or future demand growth. It is assumed that the demand at each origin cell follows a certain stochastic distribution. To represent the joint stochastic distribution of traffic demands, a set of scenarios $\Omega = \{1, 2, 3, \dots, K\}$ is introduced. A typical scenario consists of demand realizations at all origin cells. More specifically, a scenario is a vector $d^k = (d_1^k, d_2^k, \dots, d_r^k, \dots)^T, \forall r \in O$. For each scenario $k \in \Omega$, the probability of occurrence is p_k .

With the set of demand scenarios, it is feasible to formulate a stochastic program to minimize the mean of the total system delay across all demand scenarios. However, in practice, motorists and traffic engineers may be more concerned with worst-case scenarios where substantial delays may occur. To address such a risk-averse attitude on one hand and avoid being too conservative on the other hand, we optimize the signal-timing plan against a set of worst-case scenarios. More specifically, we minimize the expected 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) (26) and we name it as mean excess delay. See figure 4-1(a) for an illustration of the concept, in which the “loss” is the system delay. The probability density function and mass function of a continuous loss are shown in the figure. The right tail shaded region has an area size of $1 - \alpha$, which contains higher losses. And the mean excess delay is simply the mean of the losses in this area. Therefore, minimizing the mean excess delay is to minimize the delay incurred by the high-consequence scenarios. As aforementioned, Zhang and Yin (16) adopted the concept to synchronize actuated signals along arterials.

For each demand scenario w and one particular feasible signal plan (l, o, λ, g) , the total system delay can be computed, as described in the previous section. We denote the resulting delay as $L_w(l, o, \lambda, g)$. Consider all demand scenarios and order the total delay as $L_1 < L_2 < \dots < L_k$, 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 delay that is exceeded only with probability $1 - \alpha$. Consequently, the expected delay exceeding L_{k_α} is the mean excess delay, which can be computed as follows:

$$\phi_\alpha = \frac{1}{1 - \alpha} \left[\left(\sum_{k=1}^{k_\alpha} p_k - \alpha \right) L_{k_\alpha} + \sum_{k=k_\alpha+1}^K p_k L_k \right] \quad (5)$$

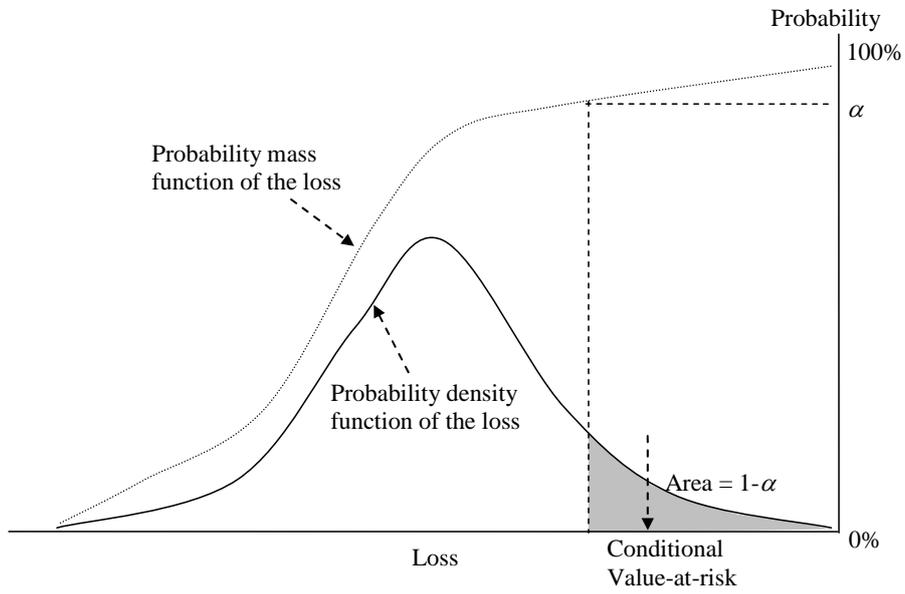
The second component in the bracket is simply to compute the mean value, and the first is to split the probability ‘atom’ at the delay point L_{k_α} to make the collective probability of scenarios considered in the bracket exactly equal to $1 - \alpha$. See figure 4-1(b) for an illustration of the concept. 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 (5) reduces to

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

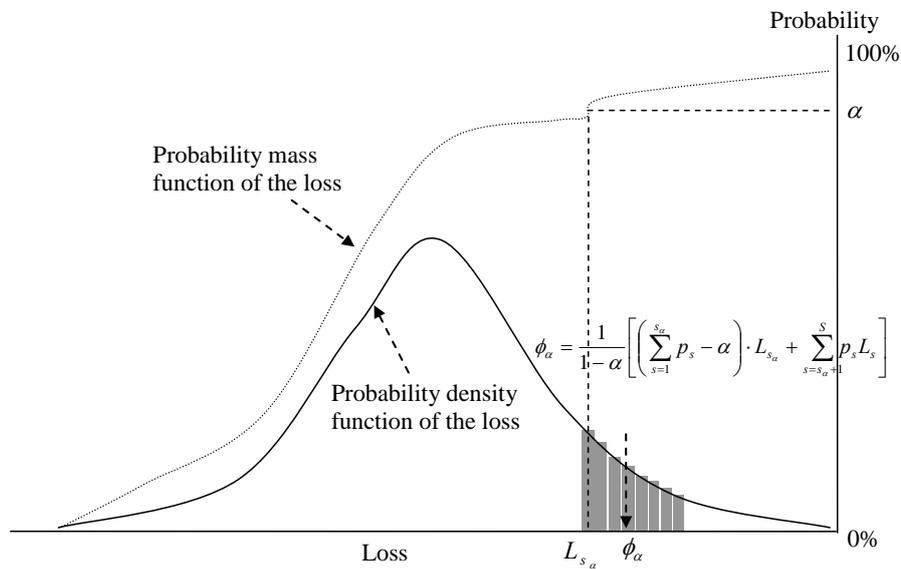
For each feasible signal plan (l, o, λ, g) , Equation (5) can be used to compute the resulting mean excess delay. Our intention is to find a signal plan that leads to the minimum value of the mean excess delay. More specifically, we attempt to minimize Equation (5) subject to multiple sets of constraints. Each set of constraints is the same as previously described for the deterministic optimization problem, but is written with respect to a specific demand scenario. Rockafellar and Uryasev (27) showed that minimizing Equation (5) is equivalent to minimizing the following equation:

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

where ξ is a free decision variable. The optimal value of the objective function is the minimum mean excess delay and the solution (l, o, λ, g) is the robust signal-timing plan.



(a) A continuous loss function



(b) A discrete loss function

Figure 4-1. Illustration of concept of mean excess delay.



5. NUMERICAL EXAMPLES

5.1 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. (24), 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 5-1 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.

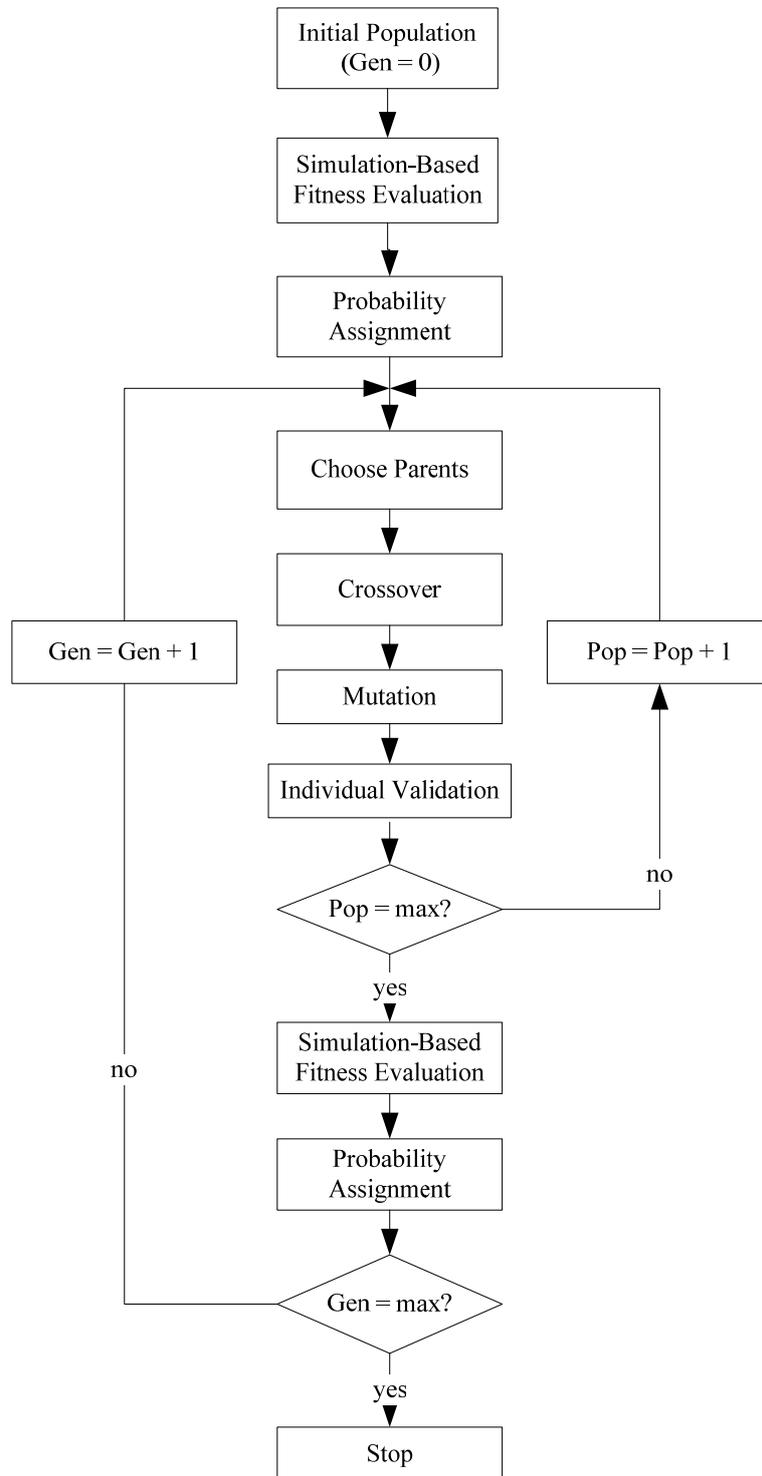


Figure 5-1. Flow chart of the simulation-based GA.



5.1.1 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 5-2 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 3-4(b)), 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.

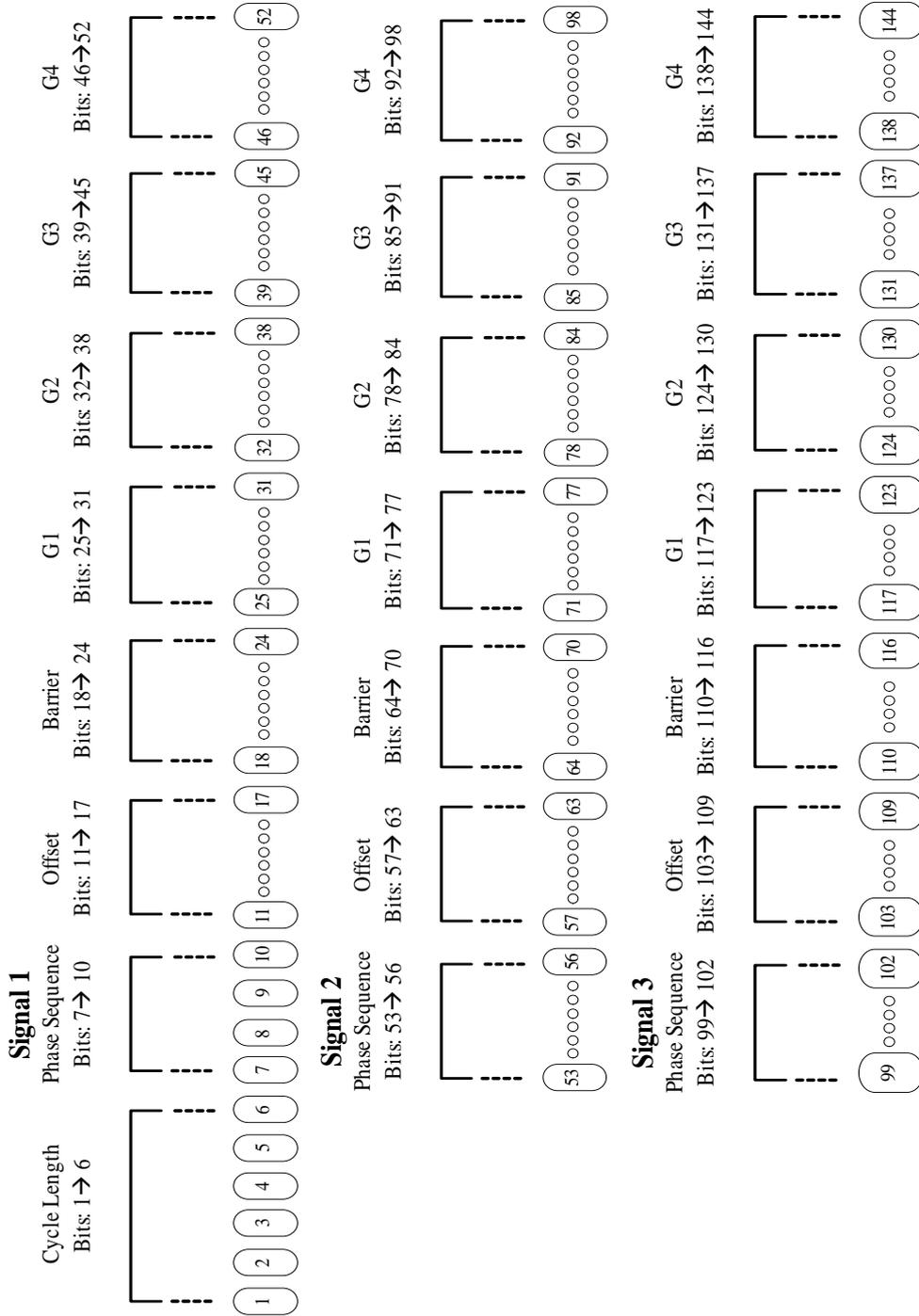


Figure 5-2. Configuration of the chromosome.



5.1.2 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 breeding the next generation.

5.1.3 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:

$$\left\{ \begin{array}{l} \frac{1}{(CVAR)^5} \\ \frac{1}{\ln(CVAR)} \end{array} \right.$$

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

5.1.4 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.

5.1.5 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%.



5.1.6 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.

5.2 NUMERICAL EXAMPLE I

5.2.1 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 5-3. 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 5-1 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.

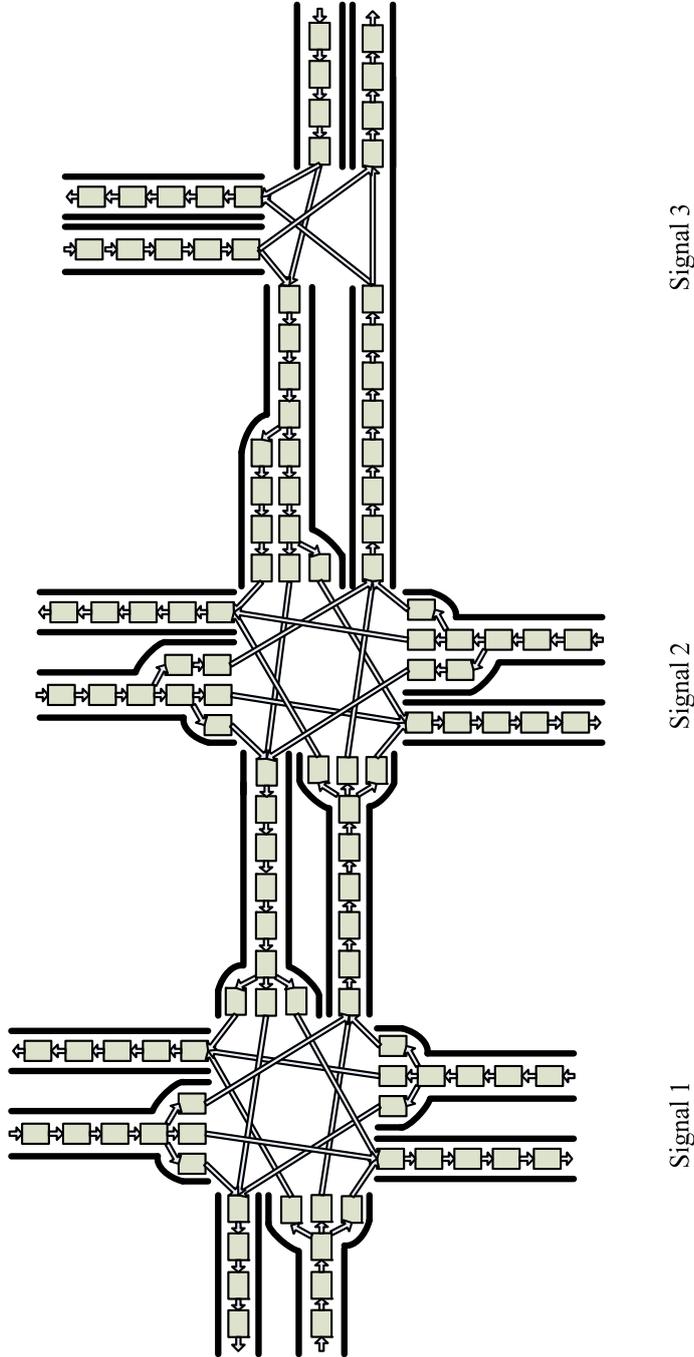


Figure 5-3. Cell representation of the three-node network



Table 5-1. Traffic Data for Three-Node Network

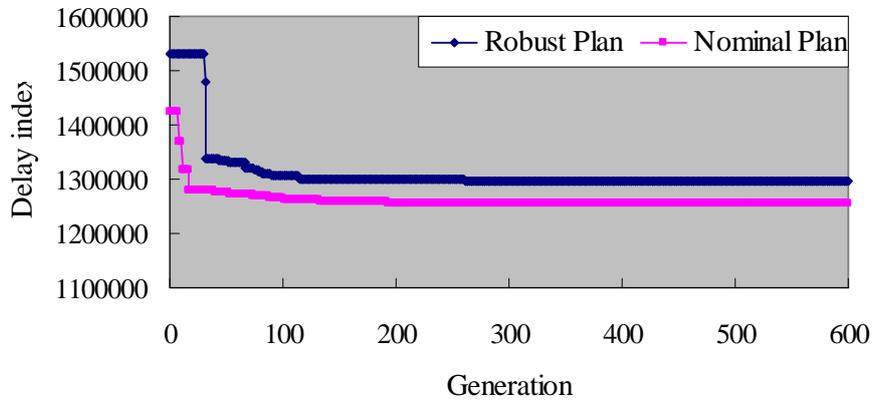
Traffic Volume		Westbound	Northbound	Eastbound	Southbound
Signal 1	Low Demand	--	162 ±27*	1223 ±180	125 ±18
	High Demand	--	462 ±45	1523 ±270	325 ±36
	Left	0.1445	0.2876	0.0291	0.6637
	Through	0.5772	0.1373	0.8960	0.2389
	Right	0.2783	0.5752	0.0750	0.0973
Signal 2	Low Demand	--	117 ±18	--	169 ±27
	High 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

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

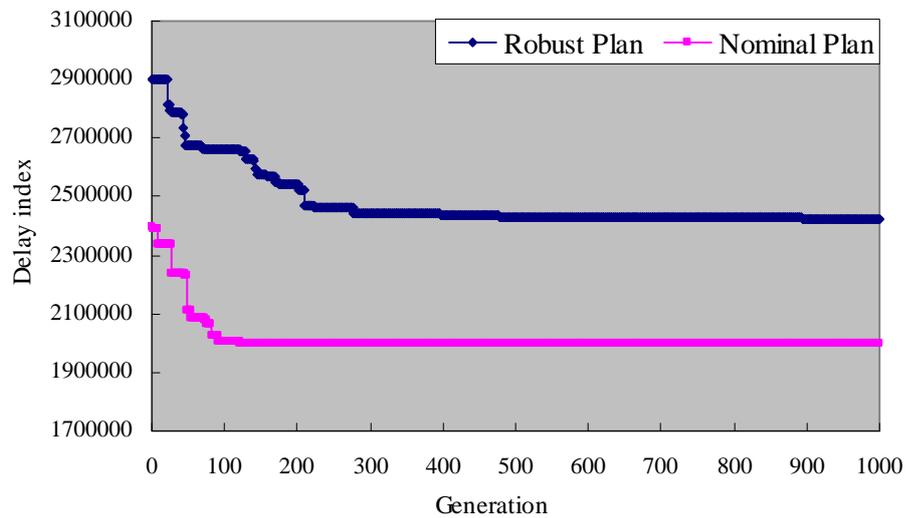
5.2.2 Plan Generation

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 5-1; the other is called nominal plan, derived by solving the enhanced deterministic signal timing model discussed in Chapter 3, with the demand in Table 5-1 without considering demand variations. 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 1,000 for the uncongested and congested case respectively. Figure 5-4 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.



(b) Uncongested case



(a) Congested case

Figure 5-4. Convergence of GA under both traffic conditions.

Table 5-2 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.



Table 5-2 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.

5.2.3 Plan Evaluation

One of the important performance measures for signal operations along arterials is the green bandwidth. For example, bandwidth maximization is the objective of the signal optimization software *MAXBAND*. For comparison, we calculate the two-way bandwidths of the signal plans presented in Table 5-2. The simplified time-space diagrams are shown in Figure 5-5 where the horizontal axis represents time and the vertical is space; the length of the solid lines means the duration of the red for the major-street traffic. The two-way bandwidths are summarized in Table 5-3.

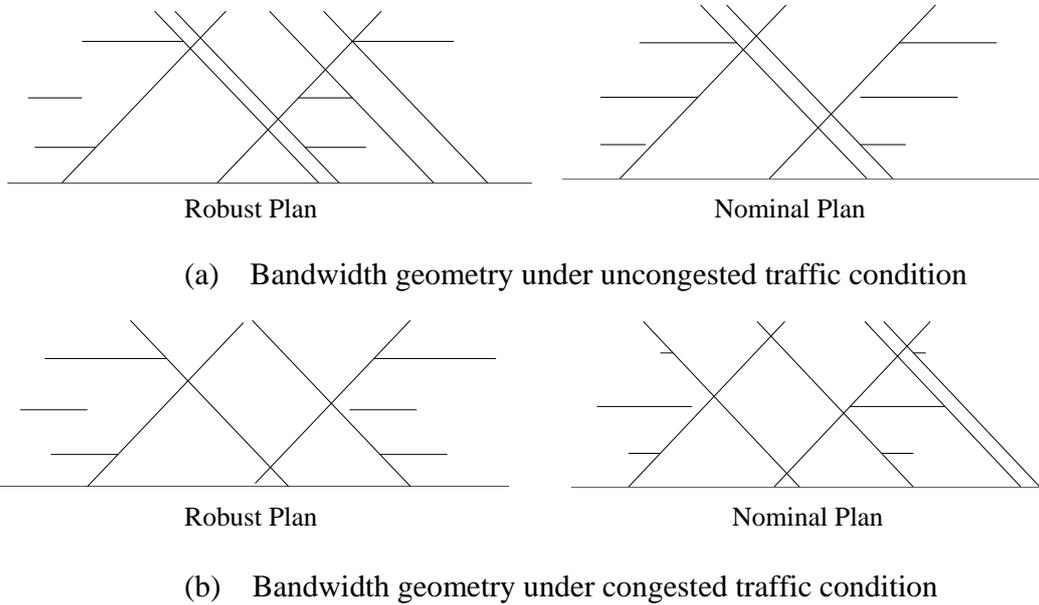


Figure 5-5. Time-space diagram of the three-node arterial.

Table 5-3 Two-Way Bandwidth Results for Three-Node Network

Uncongested Case	Cycle Length (seconds)	Two-Way Bandwidth (seconds)	Bandwidth/Cycle length
Robust Plan	80	68	0.850
Nominal Plan	80	54	0.675
Congested Case	Cycle Length (seconds)	Two-Way Bandwidth (seconds)	Bandwidth/Cycle length
Robust Plan	108	94	0.870
Nominal Plan	80	90	1.125

In the uncongested case, with the same cycle length, the robust plan provides more than 25 percent more two-way bandwidth than the nominal plan, which implies that vehicles on the major arterial will have better chance to travel without frequent stop. However, in the congested case, the nominal plan has much higher ratio of bandwidth to cycle length. This is not out of expectation, given that our formulation is mainly concerned with the control delay. When the traffic is oversaturated, it is more likely that less delay does not imply better progression and vice versa.

We further compare the robust and nominal signal plans using the microscopic CORSIM simulation, and the system delay is selected as the performance measure. Figure 5-6 is a snapshot



of the CORSIM network. In the simulation, demand scenarios are obtained by sampling the uniform distributions provided in Table 5-1. Table 5-4 summarizes the simulation result. It can be seen that the robust timing plans reduce the mean excess delay by 28.68 percent in the uncongested case and 7.46 percent in the congested case. In both cases, it also improves the average delay across all demand scenarios by over 20 percent.

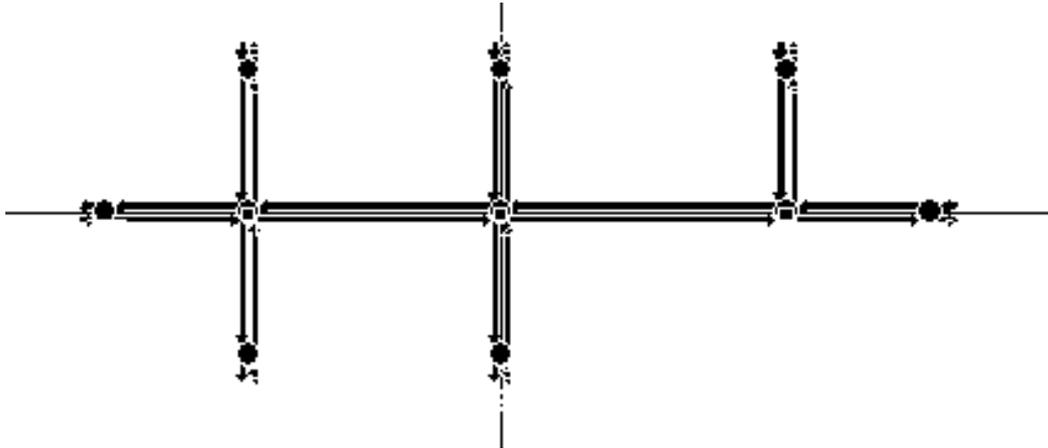


Figure 5-6. A snapshot of the three-node network in CORSIM.

Table 5-4. CORSIM Result for Three-Node Network

Taffic 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.

5.3 NUMERICAL EXAMPLE II

5.3.1 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 Road to 5th Avenue Figure 5-7 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 5-5 provides a summary of the flow data and the turning proportions at the intersections.

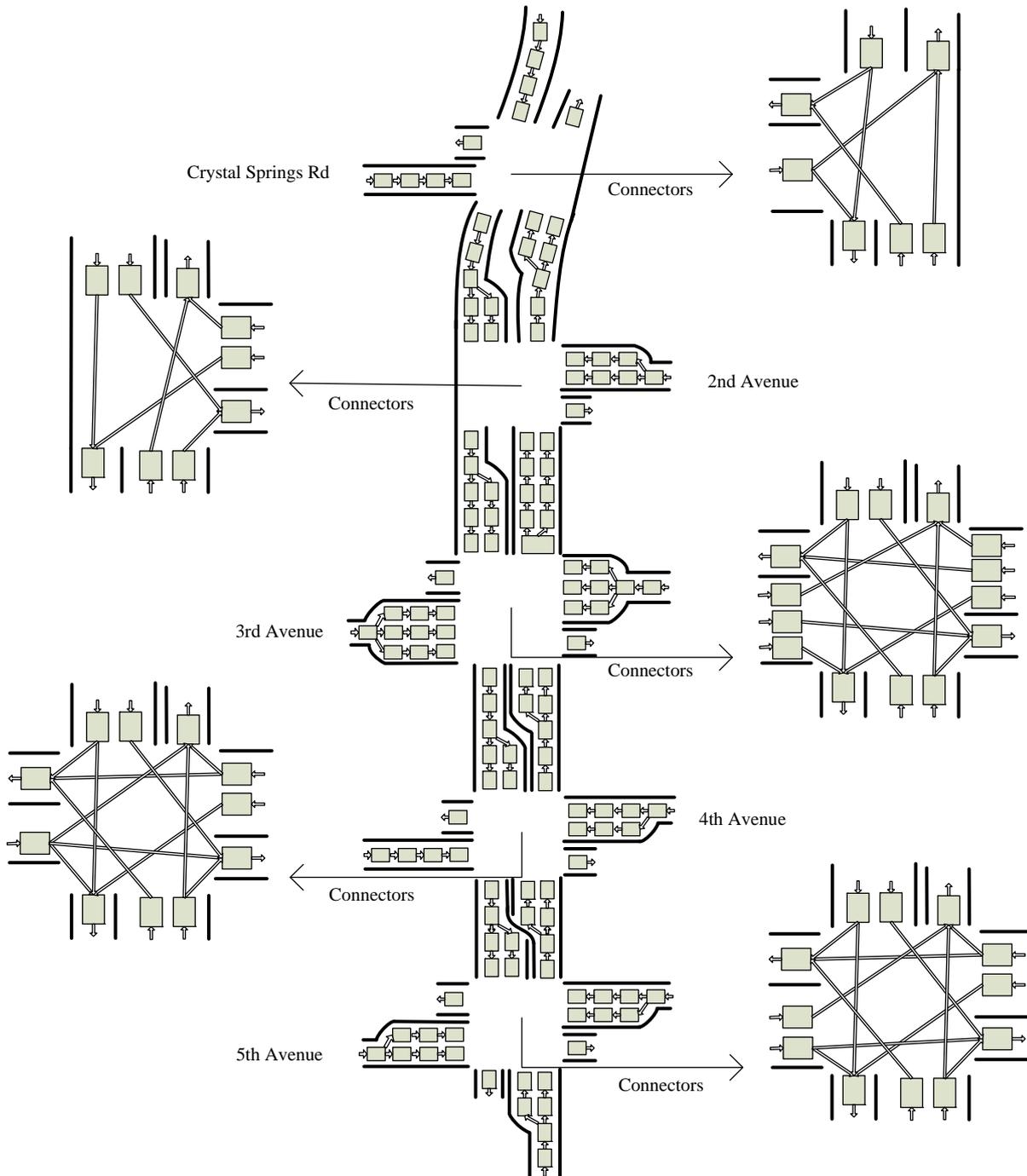


Figure 5-7. Cell representation of El Camino Real arterial.



Table 5-5. 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

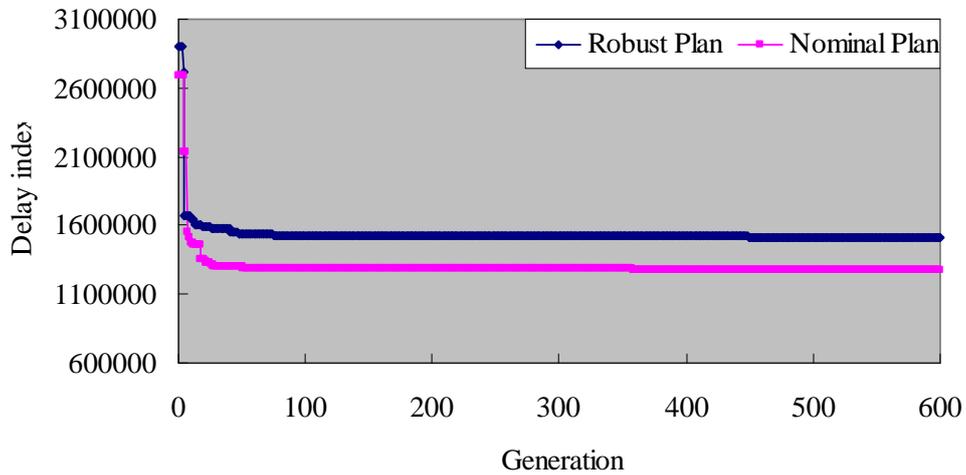
*: -- means data not applicable or available.

5.3.2 Plan Generation

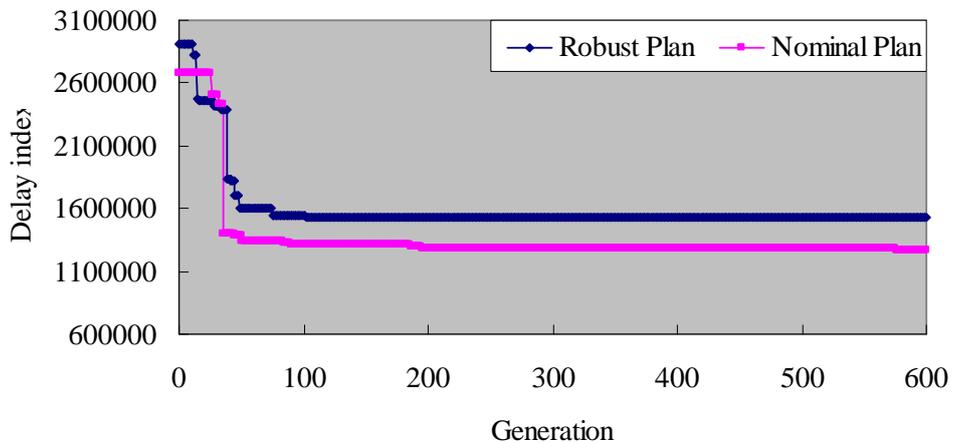
The observed flow rates are used directly as demand scenarios with equal probability of occurrence, to generate the robust plans by solving the stochastic programming model using the simulation-based GA approach. For comparison, a nominal plan is generated by solving the enhanced deterministic model with the mean demands presented in Table 5-5. Both robust and nominal plans are generated after 600 generations. Figure 5-8 shows the convergence of the GA



with both 5th-form and log-form fitness functions. It can be observed the 5th-form fitness function converges faster than the log-form counterpart.



(a) 5th-form fitness function



(b) Log-form fitness function

Figure 5-8. Convergence of GA with both fitness functions.

Table 5-6 presents the signal plans generated from both fitness functions, and their performances will be compared next. The minimum green for each phase is set as eight seconds.



Table 5-6. 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	--



5.3.3 Plan Evaluation

The comparison is also conducted via microscopic simulation with demand profiles randomly generated based on Table 5-5 assuming truncated normal distributions. Figure 5-9 is a snapshot of the CORSIM network for the corridor. Table 5-7 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 percent and 17.82 percent, and the mean excess delay reduced by 22.80 percent and 17.34 percent. 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.

Table 5-7. 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%

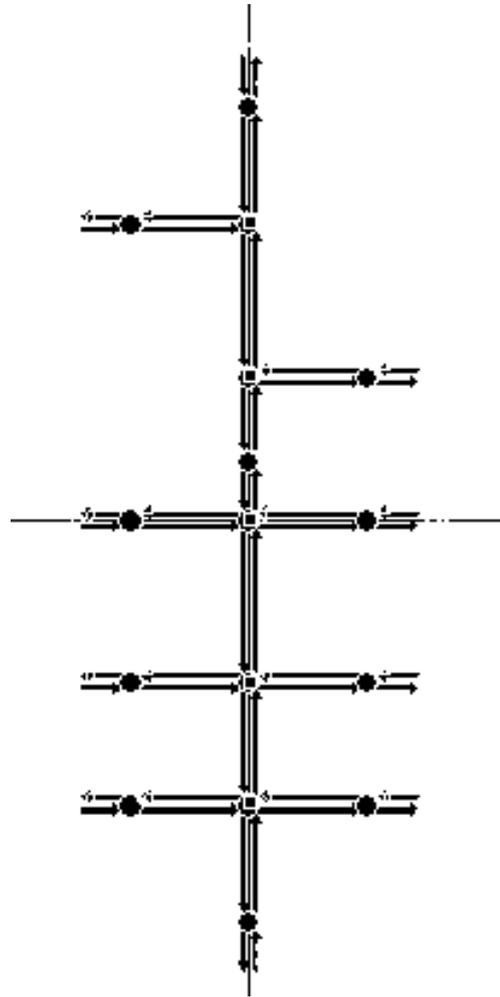


Figure 5-9. A Snapshot of the EL Camino Real Arterial in CORSIM.



6. CONCLUDING REMARKS

This report has presented a robust approach 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. Although the robust signal timing approach is applicable more widely, this report has been focused on timing models for pre-timed arterials.

Our future study may expand the proposed approach for more sophisticated corridors and grid networks. Moreover, the proposed approach can be expanded to consider other types of uncertainty, e.g., the stochastic road capacities. By specifying a number of scenarios as realizations of certain road capacities and uncertain demands simultaneously, another scenario-based stochastic program can be formulated to minimize the mean excess delay to obtain a robust timing plan.



REFERENCE

1. Smith, B. L., W. T. Scherer, T. A. Hauser, and B. B. Park. Data-Driven Methodology for Signal Timing Plan Development: a Computational Approach. *Computer-Aided Civil and Infrastructure Engineering*, Vol.17, 2002, pp.387-395.
2. Webster, F. V. *Traffic Signal Settings*. Road Research Technical Paper, No. 39, Her Majesty's Stationary Office, London, U.K., 1958.
3. Wallace, C. E., K. G. Courage, M. A. Hadi, and A. G. Gan. *TRANSYT-7F User's Guide*. University of Florida, Gainesville, FL, 1998.
4. Heydecker, B. Uncertainty and Variability in Traffic Signal Calculations. *Transportation Research*, Part B, Vol.21, 1987, pp.79-85.
5. Yin, Y. Robust Optimal Traffic Signal Timing. *Transportation Research*, Part B, 42, 2008, pp. 911-924.
6. Reiss, R.D and M. Thomas. *Statistical Analysis of Extreme Values*. 3rd ed., Birkhauser, 2007.
7. Luyanda, F., D. Gettman, L. Head, S. Shelby, D. Bullock, and P., Mirchandani. ACS-Lite Algorithmic Architecture: Applying Adaptive Control System Technology to Closed-Loop Traffic Signal Control Systems. Design Guidelines for Deploying Closed-Loop Systems. In *Transportation Research Record No. 1856*, TRB, National Academies Council, Washington, D.C., 2003, pp. 175-184.
8. Gettman, D., S. G. Shelby, L. Head, D. M. Bullock, and N. Soyke. Data-Driven Algorithms for Real-Time Adaptive Tuning of Offsets in Coordinated Traffic Signal Systems. In *Transportation Research Record No. 2035*, TRB, National Academies Council, Washington, D.C., 2007, pp.1-9.
9. Gazis, D. and R. Potts. The Oversaturated Intersection. *Proceeding, 2nd International Symposium on Transportation and Traffic Theory*, 1963, pp. 221-237
10. Gazis, D. C. Optimum Control of a System of Oversaturated Intersections. *Operations Research*, Vol. 12, 1964, pp.815-831.
11. Robertson, D. I., and R. D. Bretherton. Optimizing Networks of Traffic Signals in Real-Time: the SCOOT Method. *IEEE Trans. Vehicular Tech*, Vol.40, 1991, pp.11-15.
12. Gartner, N. H. Development and Implementation of an Adaptive Control Strategy in a Traffic Signal Network: the Virtual-Fixed-Cycle Approach. *Proceedings, 15th International Symposium on Transportation and Traffic Theory*, 2002, pp. 137-155.



13. Ribeiro, P. C. M. Handling Traffic Fluctuation with Fixed-Time Plans Calculated by TRANSYT. *Traffic Engineering and Control*, Vol.35, 1994, pp.362-366.
14. Ben-Tal, A., and A. Nemirovski. Robust Optimization—Methodology and Applications. *Mathematical Programming*, Ser.B, Vol. 92, 2002, pp.453-480.
15. Bertsimas D., and M. Sim. Robust Discrete Optimization and Network Flows. *Mathematical Programming*, Ser.B, Vol. 98, 2003, pp.49-71.
16. Zhang, L. and Y. Yin. Robust Synchronization of Actuated Signals on Arterials. In *Transportation Research Record No. 2080*, TRB, National Academies Council, Washington, D.C., 2008, pp. 111-119.
17. Little, J. D. The Synchronization of Traffic Signals by Mixed-Integer Linear Programming. *Operations Research*, Vol. 14, 1966, pp. 568-594.
18. Curtis, E. *Signal Timing under Saturated Conditions*. FHWA Office of Transportation Management/Resource Center, U.S. Department of Transportation, 2007.
19. Daganzo, C. F. The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research*, Part B, Vol. 28, 1994, pp.269-287.
20. Daganzo, C. F. The Cell Transmission Model, Part II: Network Traffic. *Transportation Research*, Part B, Vol. 29, 1995, pp.79-93.
21. Lin, W., and C. Wang. An Enhanced 0-1 Mixed Integer LP Formulation for Traffic Signal Control. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 5, 2004, pp.238-245.
22. Lo, H. A Novel Traffic Signal Control Formulation. *Transportation Research*, Part A, Vol. 44, 1999, pp.433-448.
23. Lo, H. A Cell-Based Traffic Control Formulation: Strategies and Benefits of Dynamic Timing Plan. *Transportation Science*, Vol. 35, 2001, pp.148-164.
24. Lo, H., E. Chang, and Y. C. Chan. Dynamic Network Traffic Control. *Transportation Research*, Part A, Vol. 35, 2001, pp.721-744.
25. Pavlis, Y., and W. Recker. A 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*, Vol. 43, 2009, pp.98-116.



26. Rockafellar, R. T., and S. Uryasev. Optimization of Conditional Value-at-Risk. *Journal of Risk*, Vol. 2, 2000, pp.21-41.
27. Rockafellar, R. T., and S. Uryasev. Conditional Value-at-Risk for General Loss Distribution. *Journal of Banking and Finance*, Vol. 26, 2002, pp.1443-1471.