ACTUATED COORDINATED SIGNALIZED SYSTEM

Phase I – Oversaturated Conditions

Phase II – Cycle-by-Cycle Analysis

FINAL REPORT

November 2001

Report Budget Number KLK202

NIATT Report Number N01-20

Prepared for

U.S. Department of Transportation

University Transportation Centers Grant

Prepared by

National Institute for Advanced Transportation Technology

University of Idaho

Zaher Khatib

Ahmed Abdel-Rahim

Geoffrey Judd

Krishnakanth Jagarapu

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The role of the vehicle is more important today than ever before in history, and its increased usage has led to congestion, not only in urban metropolitan areas, but also in rural small and mid-size cities. In most areas, the surface transportation network operates at capacity or over capacity, at least during some parts of the morning and evening peak hours. Spillback is a problem that frequently occurs in actuated coordinated signalized intersections that are oversaturated. Signal timing parameters for the intersections and the arterial coordination could be a major factor in the spillback phenomenon. The overall efficiency of the intersection operation could be improved by reallocating the green times between different approaches and utilizing the “wasted green time” that some approaches might have.

Phase I of the project introduces a linear optimization approach that effectively optimizes the throughput of the system. Phase II of the project presents a cycle-by-cycle analysis of actuated coordinated signalized systems’ operation, and examines whether changing the signal parameters and coordination on a cycle-by-cycle basis could lead to more effective operation of the system.

Phase I: Oversaturated Conditions

Methodology

The strategies that were developed in this research involve using linear optimization of maximum green intervals to account for the high volumes of traffic. The process includes optimizing both the phase sequence and the phase timing parameters for each of the intersections along the arterial.

The methodology was developed and tested incrementally, by adding one intersection at a time. The first formulation optimized the signal timing parameters, but not the phasing patterns, for one intersection. The second formulation included parameters to increase the efficiency of the phasing pattern. The third formulation included multiple intersections. The fourth and final formulation was for a multiple-intersection case study provided by the Idaho Transportation Department (ITD), which was located in Idaho Falls, Idaho.

A program called *LINDO* (Linear, INteractive, and Discrete Optimizer) was used for optimization in this study. After optimization, the output for the intersections along the arterial was run through the simulation program *CORSIM* to obtain a comparison. Two simulations were performed for each intersection, pre- and post-optimized.

Conclusions and Recommendations

This methodology has not yet been tested in the field, but the simulation modeling shows an increase in the operational performance of the arterial. The average total time for the entire system in the pre-optimized condition was 91.9 seconds per vehicle, with 357 vehicles per hour. The post-optimized average total time was 65.3 seconds per vehicle, with 431 vehicles per hour. While this does not altogether eliminate the oversaturation, it does provide a more effective means of processing the vehicles through the system.

The limits of the model, in terms of maximum degree of saturation, are yet unknown. Further studies could be conducted, using several different scenarios to see what limitations the model might have.

Phase II: Cycle-by-Cycle Analysis

Introduction

Traffic signal timing parameters are typically determined based on the average traffic flow, which often causes spillback or starvation at the intersection when traffic arriving at the intersection exceeds the average flow. To overcome this limitation, these parameters need to be adjusted on a cycle-by-cycle basis according to the variation in the flow arriving at the intersection.

The objective of this phase of the study is to develop a methodology that selects a particular signal timing plan for an actuated coordinated intersection based on the fluctuations in the arrival rate for each cycle, while maintaining coordination with the adjacent intersections.

Methodology

In the proposed methodology, the congestion mechanism at each approach can be described quantitatively based on the volume arriving at different intervals of the cycle. The actuated signal control parameters are then determined sequentially on cycle-by-cycle basis.

The signal control algorithm used in this research was designed to prevent spillback at different links by controlling the queue length and to minimize the total delay at the intersection. The computation was performed on a cycle-by cycle basis using the delay estimates for vehicles arriving at different intervals thought the cycle.

The model was simulated twice, using *CORSIM* and a hardware-in-the-loop simulation incorporating NIATT's new Controller Interface Device.

Conclusions and Recommendations

In the case study examined in this research, the arterial performance showed improvement over coordinated actuated signal systems. The average delay for the arterial was reduced by 8.23 percent. The average delay for the minor traffic was increased by an average of 11.1 percent, but the overall average delay for the intersection was reduced by 6.1 percent.

The limits of the model and its applicability in the field are yet unknown. Further studies could be conducted, using different traffic volumes and network configurations, to see limitations that the model might have or to more accurately assess its delay reduction potential.

Background to this Project

The role of the vehicle is more important today than ever before in history, and its increased usage has led to congestion, not only in urban metropolitan areas, but also in rural small and mid-size cities. In most areas, the surface transportation network operates at capacity or over capacity, at least during some parts of the morning and evening peak hours. Oversaturated operation of a system is the point at which the maximum hourly vehicle arrival rate is greater than the maximum hourly rate of departure. An intersection operating in oversaturated conditions will typically have long queue lengths that may not dissipate during the green phase of the signal, causing delays. These queue backups may also reduce the capacity of the system upstream from the queues.

Actuated coordinated signalized systems have become the choice of many transportation engineers, due to their ability to adjust to variation in demand. The parameters that define the operation of an actuated signal are: offset, force-off, permissive period, and yield point, which will allow the signal to return to the coordinated phases along the arterial. Signals in an actuated coordinated system must all operate under the same background cycle length. In order to achieve optimal traffic flow, all of the above parameters should be incorporated simultaneously.

If the critical degree of saturation for one intersection on the system approaches the oversaturation limit, then the whole system will be affected due to the spillback from the oversaturated intersection. This phenomenon is commonly referred to as gridlock. The spillback of the left turn traffic could lead to a blockage of the through lanes, preventing the through traffic from proceeding through the intersection. This phenomenon is known as starvation. Figures 1 and 2 show the spillback on an arterial and the starvation caused by left-turn spillback.

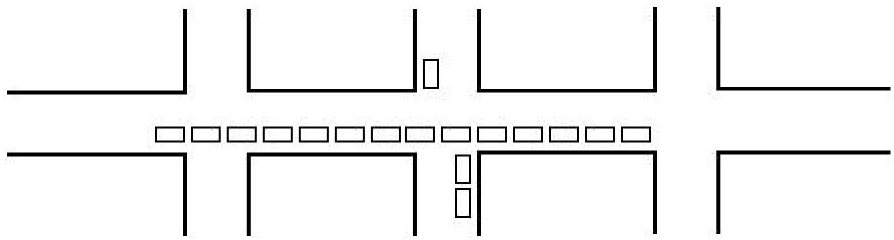

Figure 1 Spillback on the Arterial

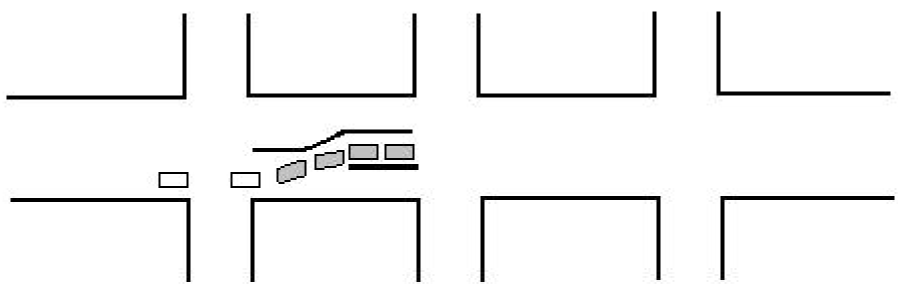

Figure 2 Starvation caused by Left-Turn Spillback

There are two problems that might cause the spillback and starvation at an intersection. The first is that the intersections along an arterial are so closely spaced that even normal amounts of traffic generate queues that can build up to the end of the block during the red phase. This is a geometric problem, and not easily corrected by signal timing. It can, however, be alleviated by running the intersections as a pair, and allowing the traffic to proceed as if it were a single intersection.

The second problem deals with the signal timing parameters for the intersections and the arterial coordination. Signal-timing parameters could be a major factor in the spillback phenomenon.
If the green time doesn't allow for enough traffic to proceed through the intersection during the green phase, or the offset between the intersections is not set properly, the queue will continue to grow beyond the link storage capacity. The overall efficiency of the intersection operation could

be improved by reallocating the green times between different approaches and utilizing the "wasted green time" that some approaches might have.

Overview of the Project—Phase I and Phase II

This research project presented had the goal of optimizing the throughput and eliminating the spillback effect at intersections controlled by actuated coordinated signalized systems. Phase I of the project introduces a linear optimization approach that effectively optimizes the throughput of the system. Phase II of the project presents a cycle-by-cycle analysis of actuated coordinated signalized systems' operation, and examines whether changing the signal parameters and coordination on a cycle-by-cycle basis could lead to more effective operation of the system.

Current Methodologies for Analyzing Oversaturated Intersections

Some of the current methods used to analyze oversaturated intersections are based upon the equations used for undersaturated conditions and are not appropriate for the oversaturated cases. For example, one of the existing methods involves calculating the control delay of the intersection and then manipulating the traffic control parameters to efficiently operate the signalized intersection on a trial iteration basis. Another common method is the use of a software program, such as TRANSYT-7F or PASSER II-90, to coordinate the oversaturated signals. These tools are developed for undersaturated conditions only and are not appropriate for oversaturated conditions. The results obtained from these programs provide a design that may lead to inefficient operation of the arterial, which may lengthen the duration of the congestion along the arterial.

An alternative method for handling oversaturation involves using approaches such as traffic demand management (TDM), in which regulations permit or prohibit certain movements in the system during the oversaturated periods. An example of TDM would be to prohibit left-turning movements during the peak periods on the arterial at every other intersection. The drawback to TDM schemes is that prohibiting certain movements can cause a loss of serviceability on the arterial.

Kuzbari [1] has investigated the oversaturation problem for an isolated intersection with the assumption that there are only two movements in an oversaturated condition. He uses some queue management strategies to determine the green times for the movements. He uses a fixed phasing scheme, which is the standard two-phase (north-south movements followed by east-west movements) as his control strategy. He points out that the typical traffic counts are taken from the discharge side of the approach, and do not truly reflect the actual demand of the intersection, only the service for that intersection. The principle behind his formulation is that, by maximizing the departure rate for the oversaturated movements, the total vehicles in queue, the length of oversaturation, and the queued delay will all be minimized.

Kim and Messer [2] discussed the oversaturation issue from the standpoint of queue management strategies. These strategies require that the signal timing change once a set maximum back of queue is obtained, in order to prevent spillback. The goal of these strategies is also to maximize the vehicle discharge from the system. The objective function, though, only focused on the external links of the system, since the intersections are very closely spaced.

Abu-Lebdeh and Benekohal [3] have generated a formulation for multiple oversaturated intersections along an arterial. The main goal for the formulation was to control the queue formation and dissipation along the arterial. The methodology was also concerned with the spillback effect on upstream signals. They discussed the characteristics of the oversaturated flow condition, which are that the flows are not steady state, the queue buildup increases the upstream departure headway, and the large queues formed create a de facto red condition for the remaining traffic. Once again, the idea was to maximize the throughput of the system by changing the traffic control parameters.

Rouphail and Khatib [4] examined a technique to optimize the phasing pattern. This optimized phasing pattern was based upon the arrival volume for each movement, and has a basic scheme that was adopted to provide the most efficient method of moving traffic. The linear formulation provides the timing for each of the green movements, to create an optimal phasing sequence for the oversaturated condition.

Overview of this Report

The report is divided into two parts. The first part begins with an introduction to Phase I of the project followed by the methodology and linear formulation employed. The results of Phase I analysis follow, along with the conclusions and recommendations. The second part of the report begins with an introduction to Phase II, followed by the formulation and methodology used in the analysis. The development of the hardware-in-the loop simulation used in Phase II is also presented, followed by the results, conclusions and recommendations.

PHASE I: OVERSATURATED CONDITIONS LINEAR OPTIMIZATION OF ACTUATED COORDINATED SIGNAL SYSTEMS

INTRODUCTION

When a signalized intersection is oversaturated, the goal is to allow the higher-volume approaches to discharge more traffic than the lower-volume approaches, so that the intersection can return to a normal operating condition as quickly as possible. The objective of this research is to generate strategies that allow this discharge to occur by adjusting signal timing parameters. The strategies that were developed in this research involve using linear optimization of maximum green intervals to account for the high volumes of traffic. The process includes optimizing both the phase sequence and the phase timing parameters for each of the intersections along the arterial.

METHODOLOGY

The methodology used in this project was developed and tested incrementally, by adding one intersection at a time. This approach was adopted because it is simpler to implement the constraints required. After the first intersection was optimized, additional intersections were added into the arterial system. The first formulation optimized the signal timing parameters, but not the phasing patterns, for one intersection. The second formulation included parameters to increase the efficiency of the phasing pattern. The third formulation included multiple intersections. The fourth and final formulation was for a multiple-intersection case study located in Idaho Falls, Idaho, which was provided by the Idaho Transportation Department (ITD).

To verify the linearly-optimized signal timing scheme, simulations of the intersections were developed using *CORSIM*, a microscopic-based program that produces several important output results such as delay and throughput volumes. The comparisons between pre-optimized operation and post-optimized operation provide a basis to support the validity of this work.

The final formulation, developed for a multiple intersection case, can be applied to any oversaturated system by inputting the geometry, traffic, and signalization parameters and optimizing the oversaturated movements. Analysis of the optimization results showed the effects of these strategies. Once these strategies are applied to the intersection signal control strategies, we expect to show a decrease in the oversaturation, and therefore a decrease in the delay per vehicle and an increase in the amount of traffic flowing through a signalized arterial.

The proposed linear formulation contains three levels of detail required in the model: System, Zonal, and Local. The system-level items are those that pertain to the entire system and are common throughout the system. Items in this level include the number of intersections, the throughput, the saturation headways, the possible background cycle lengths and the minimum green and clearance intervals. The zonal level pertains to pairs of intersections that are adjacent to each other. Items in this level include the offsets, queue lengths, storage capacity, link and block lengths, and the number of lanes. The final level is the local level, which pertains to each individual intersection. This level includes such items as the lane designations, number of approaches, arrival demand, discharge volume, green splits, and the phasing sequence and timings. The geometric, traffic, and signalization parameters used in the methodology are presented in Figures 3 though 5.

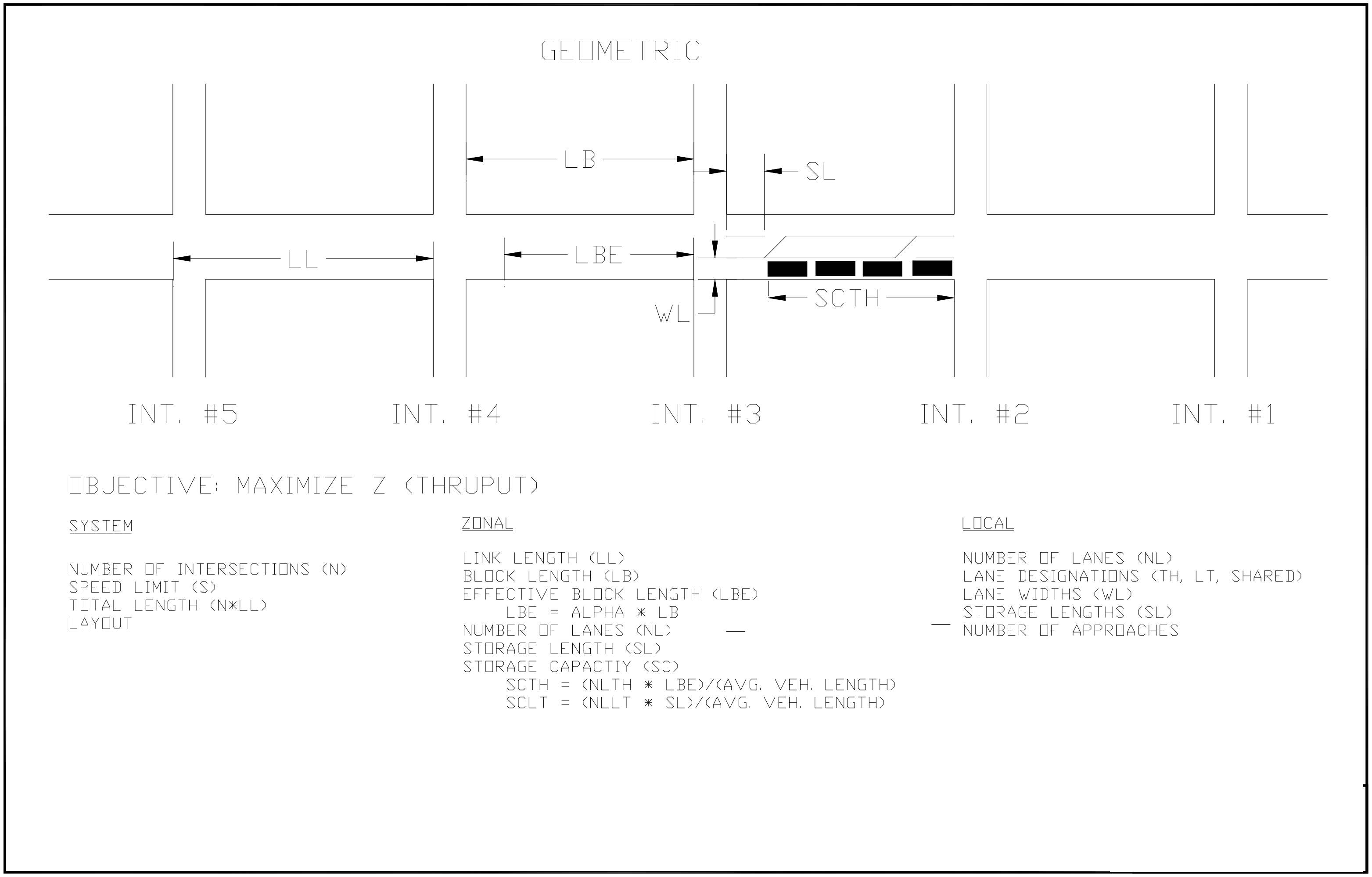

Figure 3 Geometric Parameters

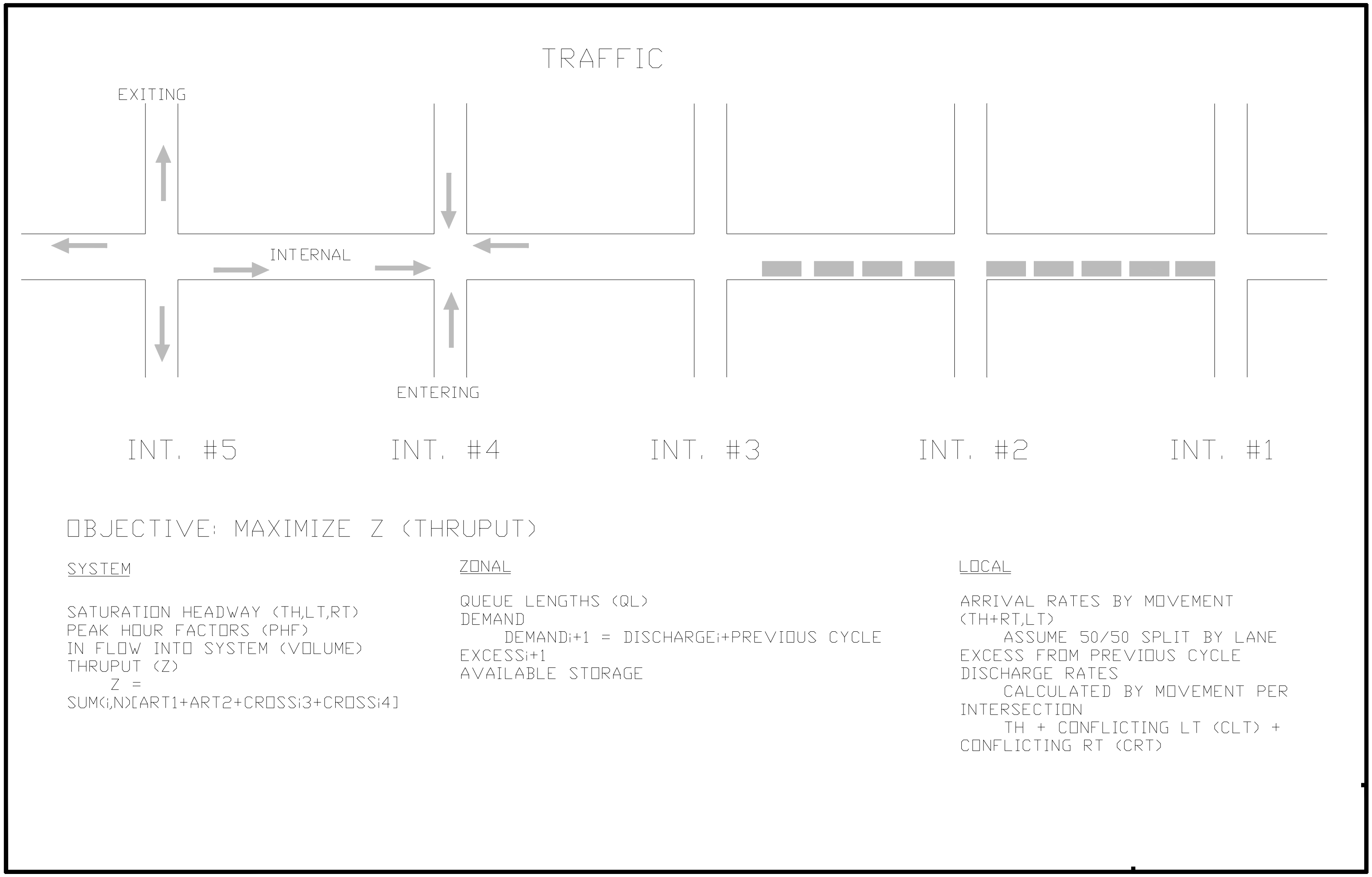

Figure 4 Traffic Parameters

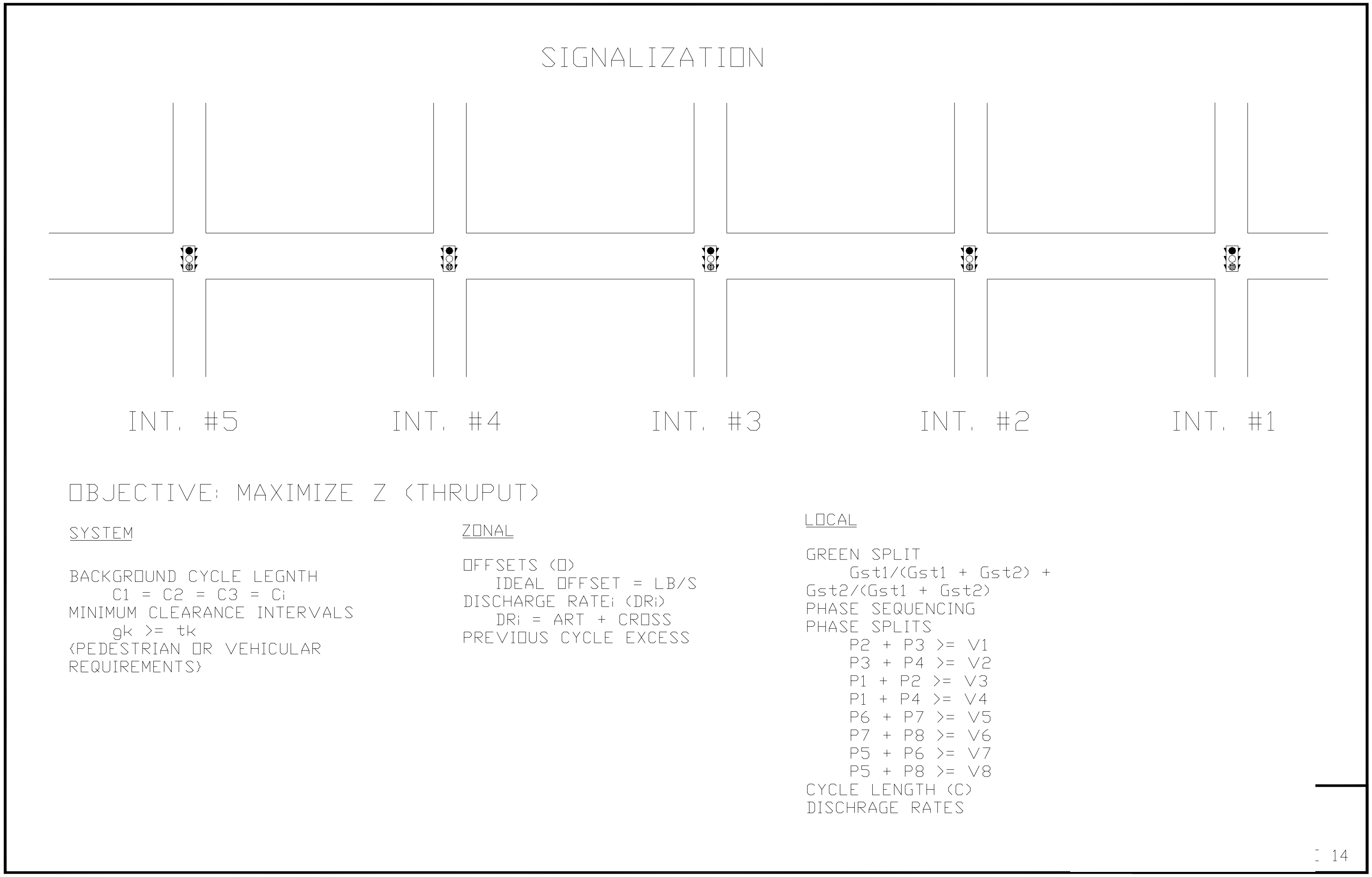

Figure 5 Signalization Parameters

Linear Programming Formulation

A program called *LINDO* (Linear, INteractive, and Discrete Optimizer) [5] [6] was used for optimization in this study. The objective function of the linear formulation is to maximize the number of vehicles leaving the system, and thereby reduce the congestion within the system. This is accomplished by multiplying the inverse of the average saturation headway (h) of the vehicles in the system by the green time (G_{ij}) for each of the movements at the signalized intersections. The average headway values range between 1.8 and 2.2 seconds. A headway value of 2.0 seconds was used for the case study so that there would be no shift in the oversaturated movements. The shift might occur when one of the oversaturated movements is optimized, since it decreases the available time for other movements and might change the other movements into an oversaturated movement. The green time is optimized for each movement. The general form for this objective function is shown below

$$MAX \quad (\sum \frac{1}{h} * G_{ij})$$

Where i is the intersection number and j is the movement number. The i value has a range of 1 to N, the number of intersections, and the j value has a range from 1 to the number of movements at the intersection, maximum of eight. The movement numbers are based upon the NEMA standard movement numbering system.

Determine Oversaturated Movements

The first step in the model is to determine the movements that are oversaturated for the intersection. This is performed outside of the linear program using a software program such as *Highway Capacity Software HCS* or by manual computations. This is an important step in the complete formulation, because one must know which of the movements are oversaturated in order to allocate the green times properly. There are two drawbacks to performing this pre-computation.

One drawback is that after the signal timings have been changed the capacity will change, because the capacity is dependant upon the green-to-cycle-length ratio of the approaches. The second drawback is that this computation takes place outside of the linear program, and therefore requires a little more time and effort. The idea is that even though these values change with the optimization, the overall capacity will increase enough to reduce the severity of the congestion.

The first set of constraints in the model are used to assign the green for each movement based upon the arrival rates. This is done by relating a queue-growing speed function for a movement to the storage length for that movement at the intersection. This queue-growing function consists of the green times for that movement (G_{ij}), the cycle length (C_i), the arrival rate (V_{ij}), and the saturation flow rate (S_{ij}). The equation for the queue-growing function is shown below.

$$T_{ij} = V_{ij} * C_i - G_{ij} * S_{ij}$$

The green is split between the left-turning movements and through movements. The storage length (L_{ij}) on the denominator of the equations below represents either the storage length, for left-turn bays, or the effective block length for the through movements. There are two sets of equations to define the green times, shown below.

$$\frac{T_{i1}}{L_{i1}} = \frac{T_{i3}}{L_{i3}} = \frac{T_{i5}}{L_{i5}} = \frac{T_{i7}}{L_{i7}}$$

$$\frac{T_{i2}}{L_{i2}} = \frac{T_{i4}}{L_{i4}} = \frac{T_{i6}}{L_{i6}} = \frac{T_{i8}}{L_{i8}}$$

These equations lead to six different constraints to the model, each having the form shown below. This is done for each oversaturated movement at each intersection. The reason that only oversaturated movements are used in this equation is that the equation only defines the green time for those movements based upon the queue growth during the cycle. An example of one of the constraints is shown below.

$$T_{i4} * L_{i2} - T_{i2} * L_{i4} = 0$$

Other Critical Movements

The next step is to find other critical movements (not saturated) at the intersection. In this step, binary variables are used to develop decision logic in the equation. These binary variables denote which movements are critical, and when optimized, the critical movements are denoted by a value of one or zero, depending upon which of the movements is critical. The binary variable is the w_{ij} variable, with the lower case *ij* indicating the intersection and movement of interest. There are only four critical movements per intersection, as shown in Figure 6 below. The selection is between the movements in each of the four boxes.

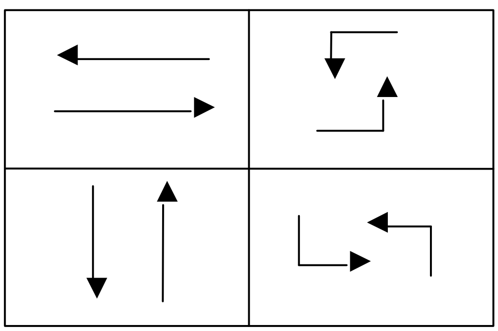

Figure 6 Critical Movement Selections

If the w_{ij} is equal to 1, then the first movement, the top or left-hand side movement, is a critical movement. If the w_{ij} is equal to 0, then the second movement is critical. There are two equations to constrain the binary variable and allow for the selection logic to properly choose the correct solution. From the linear formulation, the two equations are shown below:

$$V_{i6} * w_{i1} \leq V_{i2}$$

$$V_{i2} * w_{i1} \geq V_{i2} - V_{i6}$$

Phase Selection

Next is the phase selection. In this step, the phases are selected based on the critical *v/s* ratios. Below are the equations for the phase selections.

Dual Left for East-West

$$PHF * P_{i1} + PHF * P_{12} >= V_{i5} / S_{i5}$$
$$PHF * P_{i1} + PHF * P_{i4} >= V_{i1} / S_{i1}$$

Dual Left for North-South

$$PHF * P_{i5} + PHF * P_{i6} >= V_{i3} / S_{i3}$$
$$PHF * P_{i5} + PHF * P_{i8} >= V_{i7} / S_{i7}$$

Thru Phases for East-West

$$PHF * P_{i2} + PHF * P_{i3} >= V_{i2} / S_{i2}$$
$$PHF * P_{i3} + PHF * P_{i4} >= V_{i6} / S_{i6}$$

Thru phases for North-South

$$PHF * P_{i6} + PHF * P_{i7} >= V_{i8} / S_{i8}$$
$$PHF * P_{i7} + PHF * P_{i8} >= V_{i4} / S_{i4}$$

Where *PHF* is the peak hour factor and P_{ij} is the phase split.

Discharge Rates

The next thing to be formulated is the three discharge rates. The discharge rate is a zonal constraint, between intersections. These three constraints are used for internal queue management. The idea is that the red time for a phase should be short enough to accommodate the storage capacity for either intersection approach.

The first constraint takes into account the downstream storage capacity, using a weighted average of the storage capacities. This constraint is used as a means to control the discharge of the traffic at the intersection of interest and limits the queue that would develop at the downstream intersection. The total number of vehicles in this discharge calculation is determined through several factors that are held as constants. An example of the equation for the discharge rate is shown below,

$$C_i - G_{i6} \leq \frac{V_{(i+1)6}}{V_{(i+1)6} + V_{(i+1)1}} * \frac{SC_{TH(i+1)}}{h} + \frac{V_{(i+1)1}}{V_{(i+1)6} + V_{(i+1)1}} * \frac{SC_{LT(i+1)}}{h}$$

where V_{ij} is the arrival rate in vehicles per second, C_i is the cycle length, G_{ij} is the green time, h is the saturation headway, and SC_{ij} is the storage capacity for the movement in number of vehicles. This formulation accounts for the accumulation of vehicles in the SC_{ij} storage area, and allows these vehicles to release from the intersection before setting the discharge rate to the arrival rate. The SC_{ij} term is an input parameter based upon the geometry of the system, computed by the equation below.

$$SC_{TH(i)} = \frac{N * (LB) * (a)}{Avg.veh.Length} \quad For\, through\, movements$$

$$SC_{LT(i)} = \frac{N * (SL)}{Avg.veh.Length} \quad For\, left - turn\, movements$$

Where LB is the block length, α is a ratio to account for an effective block length, N is the number of lanes, and SL is the storage length.

The second constraint was used to control the internal queues, which accounts for the storage capacity and the initial queue at the intersection of interest. This constraint considers the storage capacity of each movement and the departure rate for that movement. It sets a lower bound on the discharge rate required to serve as much of the existing queue as possible. An example of the constraint is shown below.

$$C_i - G_{ij} \geq \frac{1}{h} * SC_{ij}$$

The third constraint is used to further restrict the internal queues, based upon the discharge rates for each of the movements. This constraint is used only for the arterial, since it only concerns the amount of traffic that is on the links between the two intersections. The equation is based upon the idea of input minus output equals the storage. The equation is shown below.

$$\frac{G_{i6}}{V_{i6}} + \frac{G_{i3}}{V_{i3}} - \frac{G_{(i+1)6}}{V_{(i+1)6}} - \frac{G_{(i+1)1}}{V_{(i+1)1}} \leq SC_{TH(i)}$$

The above constraints are used to make certain that the internal queues are not exceeding the effective block length, which is set as some percentage of the full block length.

Define Cycle Length

The final step is to define a cycle length for the system. The sum of each of the rings in the dual ring controller defines the cycle length. In other words, there are two points along the cycle when the phases will need to be in synchronization? at the end of the major movement and the end of the minor movement. The critical, or longest period of time for these two points, will define the cycle length for any actuated traffic signal. Two summations are used to constrain the cycle length. The equations are shown below:

$$G_{i1} + G_{i2} + G_{i3} + G_{i4} - C_i \leq 0$$

$$G_{i5} + G_{i6} + G_{i7} + G_{i8} - C_i \leq 0$$

The cycle length could be the same for all intersections along the arterial or it could vary. In this research, only same cycle length is analyzed.

The minimum green times are based upon pedestrian clearances or minimum green initial intervals. The cycle length is bounded to provide a more effective operation and to reduce the amount of delay incurred by the red phase. This was done to prevent the extraordinarily high delay that would result from unreasonably long cycle length. The cycle lengths may also be coordinated along the arterial, using a common cycle length. Since the operation deals with oversaturated conditions, the assumption of similar operations will occur for each subsequent cycle.

Linear Program

The linear programming formulation used in the analysis is as follow:

FOR i INTERSECTIONS
Using the NEMA phase numbers
ijk - i = intersections; j = movements; k = phases

MAX ? $_{ij}$ (1/h * G_{ij})

SUBJECT TO

Intersection #i

Assign green to saturated movements

Left-turn movements

$$\frac{T_{i1}}{L_{i1}} = \frac{T_{i3}}{L_{i3}} = \frac{T_{i5}}{L_{i5}} = \frac{T_{i7}}{L_{i7}}$$

Through movements

$$\frac{T_{i2}}{L_{i2}} = \frac{T_{i4}}{L_{i4}} = \frac{T_{i6}}{L_{i6}} = \frac{T_{i8}}{L_{i8}}$$

Identify other critical movements

$V_{i6}\, w_{i1} <= V_{i2}$
$V_{i2}\, w_{i1} >= V_{i2} - V_{i6}$

$V_{i5}\, w_{i2} <= V_{i1}$
$V_{i1}\, w_{i2} >= V_{i1} - V_{i5}$

$V_{i4}\, w_{i3} <= V_{i8}$
$V_{i8}\, w_{i3} >= V_{i8} - V_{i4}$

$V_{i7}\, w_{i4} <= V_{i3}$
$V_{i3}\, w_{i4} >= V_{i3} - V_{i7}$

Phase Selection

Dual Left for East-West

$PHF * P_{i1} + PHF * P_{12} >= V_{i5} / S_{i5}$
$PHF * P_{i1} + PHF * P_{i4} >= V_{i1} / S_{i1}$

Dual Left for North-South

$PHF * P_{i5} + PHF * P_{i6} >= V_{i3} / S_{i3}$
$PHF * P_{i5} + PHF * P_{i8} >= V_{i7} / S_{i7}$

Thru Phases for East-West

$PHF * P_{i2} + PHF * P_{i3} >= V_{i2} / S_{i2}$
$PHF * P_{i3} + PHF * P_{i4} >= V_{i6} / S_{i6}$

Thru phases for North-South

$PHF * P_{i6} + PHF * P_{i7} >= V_{i8} / S_{i8}$
$PHF * P_{i7} + PHF * P_{i8} >= V_{i4} / S_{i4}$

Set the discharge rate for movements based upon queue downstream int.

$C_i - G_{ij} <= SL_{(i+1)j}$

Set the discharge rates for movements based upon queue at upstream int.

$C_i - G_{ij} >= SL_{ij}$

Restraints based upon internal storage capacity

$$\frac{G_{i6}}{V_{i6}} + \frac{G_{i3}}{V_{i3}} - \frac{G_{(i+1)6}}{V_{(i+1)6}} - \frac{G_{(i+1)1}}{V_{(i+1)1}} \leq SC_{TH(i)}$$

Cycle Length

$G_{i1} + G_{i2} + G_{i3} + G_{i4} - C_i <= 0$
$G_{i5} + G_{i6} + G_{i7} + G_{i8} - C_i <= 0$

Green times lower bound

$G_{ij} >= \text{Min Green}_{ij}$

Cycle length

$C_i >= \text{Min C}$
$C_i <= \text{Max C}$

Coordinate Cycle Lengths?

$C_i - C_{(i+1)} = 0$ (Optional Constraint)

END

INT w_{i1}
INT w_{i2}
INT w_{i3}
INT w_{i4}

OPTIMIZATION AND TESTING

Once the constraints were placed into the linear optimization program, they were optimized. After optimization, the output for the intersections along the arterial was used in the simulation program *CORSIM* to obtain a comparison. Two simulations were performed for each intersection, pre- and post-optimized. Several simulations were performed for each condition, using different random seeds to get a better idea of how well the model was operating.

The model was tested for one, two, three, and five intersections along the arterial. This was done to check the step-wise development of the model and to make certain that each parameter provided successively better operation along the arterial. The evaluation of the test results is provided in the next section.

EVALUATION OF OPTIMIZATION STRATEGY

With this optimization strategy, any intersection geometry and traffic data can be entered into the constraints of the model, and the model will produce an optimal signal timing strategy for those intersections. The signal timing parameters can be tested in simulation models or implemented directly into the controller in the field. By using a simulation model, however, the operation of the intersection can be evaluated with the new signal timing strategy and can also compared with the operation of the original signal timing strategy.

The results of several simulations are discussed in this section. The first runs focused upon single intersections with some of the approaches oversaturated. The next step was to add complexity by changing how the optimization was performed, or by adding other intersections along the arterial. The last step was to run the case study and simulate the results. This step-wise manner of testing the model helped to refine the model and it also simplified the development of the logic behind the optimization.

After each optimization, the base case was compared to the modified case to check the measures of effectiveness, total travel time and vehicle discharge. The simulation was performed in

CORSIM, and the output of the base case and modified case were compared. The outputs of the measures of effectiveness are discussed below.

It is important to note that the measures of effectiveness were compared on the entire system. This means that there might be some approaches in which the performance will degrade, and others in which performance will improve. The overall improvement outweighs the degradation. In Figure 7 below, a network is shown inside the gray box, and the area outside of the gray box illustrates where the measures of effectiveness were taken. Internal measurements were also taken along each of the links, but the focus was upon the vehicles leaving the system.

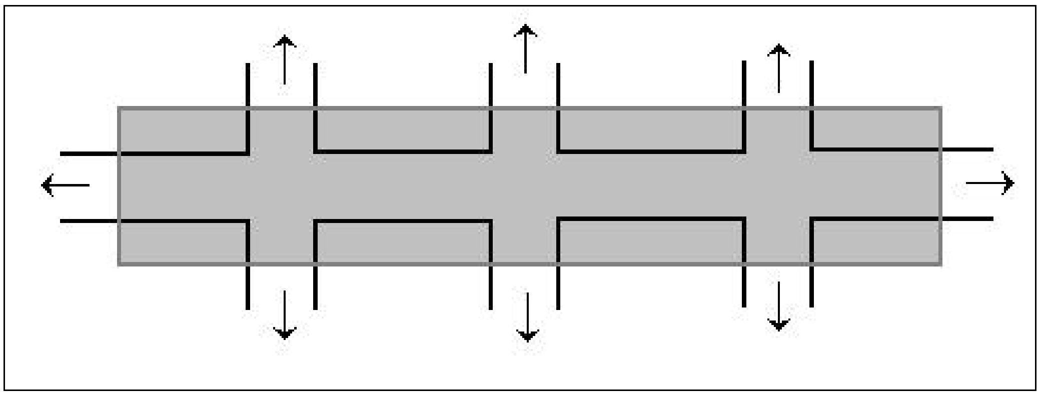

Figure 7 Network Boundaries

Optimization One – Isolated Intersection

The first step was to use the methodology on an isolated intersection with several movements in an oversaturated condition. The intersection layout shown is from the first intersection in the case study, Broadway and Yellowstone, and the volumes were increased to show the effects of oversaturation at the intersection. This intersection was chosen because it included all eight movements.

In the next steps, additional intersections from the Broadway arterial were added to the left side of this intersection. Each intersection that was added is presented with the drawing of the intersection as well as the input volumes used for the intersection. Figure 8 shows the geometry of the first intersection.

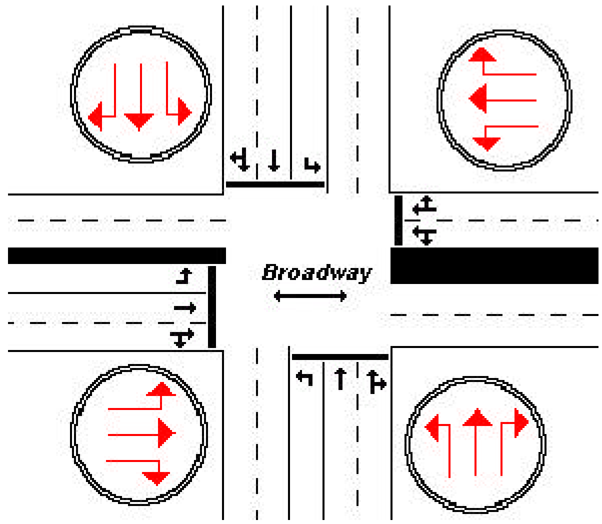

Figure 8 Geometry for Intersection #1

The intersection was simulated to show the existing conditions of the system, and then the volumes were used in the linear formulation to be optimized. Table 1 shows the change in the signalization parameters between the pre- and post-optimized signalized intersection. The signalization parameters shown in the pre-optimized phasing scheme are the requirements for the pedestrian crossing times.

Table 1 Pre- and Post-Optimized Signal Timing (One Intersection Analysis)

Phase/Movement	Pre-optimized Green	Post-optimized Green
1	16	42
2	31	40
3	26	13
4	46	60
5	16	44
6	31	16
7	26	13
8	46	16

For this first case study, the travel time and volumes were averaged for both the entering and exiting traffic. The average pre-optimized total time entering was 205.6 seconds per vehicle and the average total time exiting was 24.6 seconds per vehicle. The average post-optimized total time entering was 95.7 seconds and the average total time exiting was 25.1 seconds. The total time entering decreased by 53.4-percent, and there was a 2.0-percent increase in the total time exiting. The average pre-optimized volume entering was 564 vehicles per hour and average volume exiting was 568 vehicles per hour. The average post-optimized volume entering was 693 vehicles per hour, and the average volume exiting was 690 vehicles per hour. The volume entering increased by 22.9-percent, and the volume exiting increased by 21.5percent.

The overall system-wide performance results showed that the pre-optimized time was 115.1, and the volume was 566. The post-optimized system-wide performance results showed that the time was 60.4, and the volume was 692. The percent change for each of those was a 47.5-percent decrease in total time and a 22.3-percent increase in the volume. While these results are for only one intersection, the basis for the optimization procedure shows that there is a good foundation to perform further tests on more intersections.

Optimization Two - Two Actuated Intersections

The second formulation included two actuated intersections, which were placed 360 feet apart. There is a high probability of queue spillback at this distance, so the internal queue management of the formulation would be tested. The data were taken from the Broadway and Yellowstone and Broadway and Shoup intersections. The only difference between Optimization One and Two is that the volumes were increased to show the effect better, and some of the approaches were increased to add difficulties to the progression of the movements. In this situation there is the possibility of queue spillback or starvation, both of which were taken into account in the formulation. The map of the second intersection is shown in Figure 9.

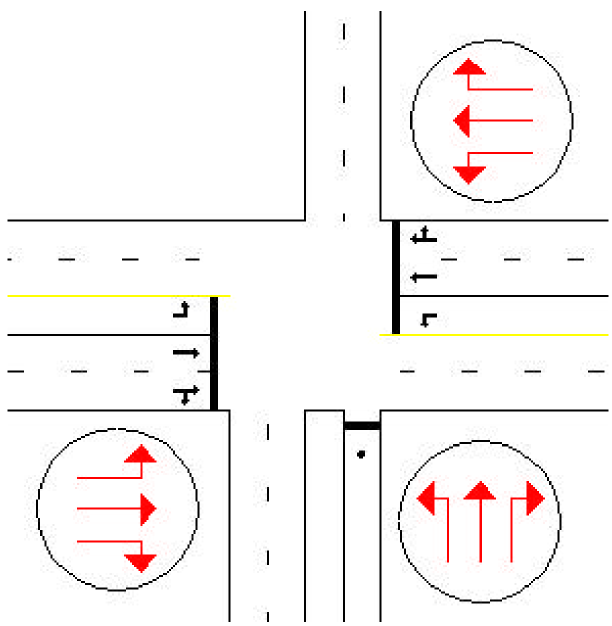

Figure 9 Geometry for Intersection #2

The signal timing parameters, both pre- and post-optimized, are shown in Table 2. The intersections are running an actuated coordinated phasing pattern, but the values listed are the values for the maximum green times. The post-optimized values are for the green times of the phases after optimization. These times were placed into the phasing pattern for the first optimization.

Table 2 Pre- and Post-optimized Signal Timing (Two Intersections Analysis)

Broadway and Yellowstone			Broadway and Shoup		
Movement	Pre-optimized	Post-optimized	Movement	Pre-optimized	Post-optimized
1	15	42	1	30	36
2	30	40	2	30	90
3	25	13	3	20	54
4	25	60	4	0	0
5	15	44	5	30	36
6	30	16	6	30	38
7	25	13	7	0	0
8	25	16	8	20	85

Once several simulations were performed, the results were averaged and compared to each other for the change in the operational characteristics of the arterial. For this paired intersection arterial, three areas were compared, the entering links, the exiting links, and the internal links.

For the pre-optimized intersections the average total time of vehicles entering the system was 312.5 seconds per vehicle and the average volume was 510 vehicles per hour. The average total time of vehicles exiting the system was 24.6 seconds per vehicle and the volume was 414 vehicles per hour. For the internal links the average total time was 91.9 seconds per vehicle and the volume was 642 vehicles per hour.

For the post-optimized intersections the average total time of the vehicles entering the system was 291.3 seconds per vehicle and the volume was 567 vehicles per hour. This was a 6.8-percent decrease in average total time, and 11.2-percent increase in the average volume. The average total time of vehicles exiting the system was 25.0 seconds per vehicle and the volume was 475 vehicles per hour. There was an increase in the average total time of 1.6-percent, and the volume increased 14.7-percent.

For the internal links, the average total time was 101.3 seconds per vehicle and the volume was 620 vehicles per hour. This was a 10.2-percent increase in average total time, and a 3.4-percent decrease in the average volume.

The degradation of the internal links’ average total time and volume created some uncertainty about what was happening in the system. There was also the fact that the internal queues were being minimized in the formulation, so that was taken into consideration.

The overall system performance between the pre- and post-optimized scenarios was then compared. The results show that for the pre-optimized scenario, total time is 145.7 seconds per vehicle and the volume is 486 vehicles per hour. The post-optimized total time is 139.2 seconds per vehicle and the volume is 533 vehicles per hour. This is a 4.5 percent decrease in the total time and a 9.7 percent increase in the average volume in the system.

Optimization Three – Three Intersections

The next step was to run a study using three intersections, Broadway and Yellowstone, Broadway and Shoup, and Broadway and Park. The results for this set of data are compiled into the system performance, since this is really the measure of the total system. The geometry of this intersection is shown in Figure 10.

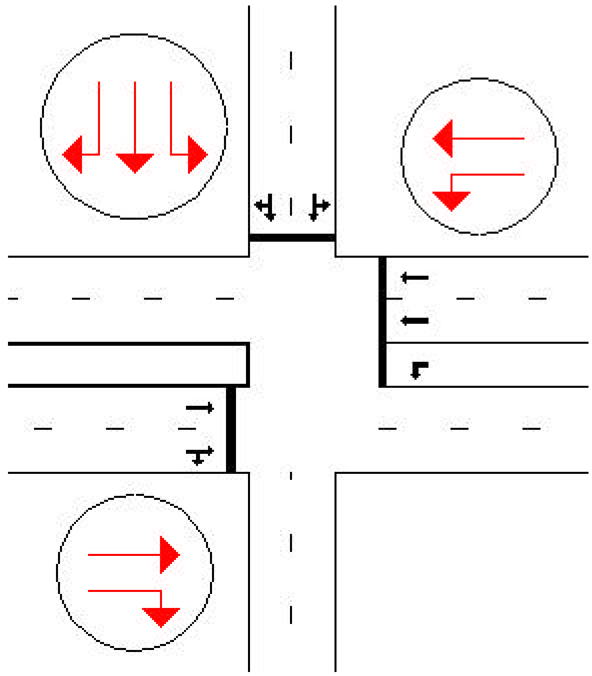

Figure 10 Geometry for Intersection #3

The signal timing parameters are shown in Table 3, both pre- and post-optimized. The intersections are running a coordinated-actuated phasing pattern; but the values listed are the values for the maximum green times. The post-optimized values are for the green times of the phases after optimization. These times were placed into the phasing pattern for the second optimization.

Table 3 Pre- and Post-optimized Signal Timing (Three Intersections Analysis)

Broadway and Yellowstone			Broadway and Shoup			Broadway and Park		
Move-ment	Pre-optimized	Post-optimized	Move-ment	Pre-optimized	Post-optimized	Move-ment	Pre-optimized	Post-optimized
1	15	42	1	30	36	1	30	11
2	30	40	2	30	90	2	30	27
3	25	13	3	20	54	3	0	0
4	25	60	4	0	0	4	20	85
5	15	44	5	30	36	5	30	11
6	30	16	6	30	38	6	30	75
7	25	13	7	0	0	7	20	85
8	25	16	8	20	85	8	0	0

The same methodology was applied to this arterial, which also showed an improvement in operational characteristics. For this three-intersection arterial, the only measure of performance was the overall system performance. The results show that for the pre-optimized scenario, total time is 110.4 seconds per vehicle and the volume is 458 vehicles per hour. The post-optimized total time is 56.7 seconds per vehicle and the volume is 606 vehicles per hour. This is a 48.6 percent decrease in the total time and a 32.3 percent increase in the average volume in the system. These overall results are the actual performance over the entire system, which is not commonly shown for arterials. In an oversaturated case though, these measures are the best representation of how the system is performing over the time period.

Optimization Four – Case Study

The last step was to run the case study from Idaho Falls, Idaho. This involved five oversaturated intersections, which are the most critical intersections on the arterial. The geometries for the last two intersections are shown in the Figures 11 and 12.

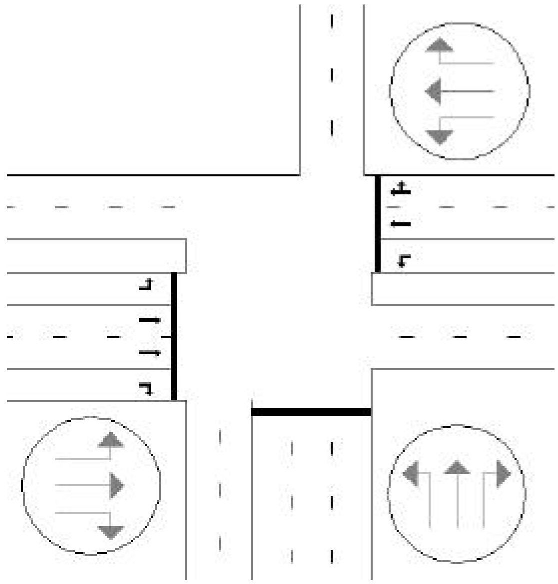

Figure 11Geometry for Intersection #4

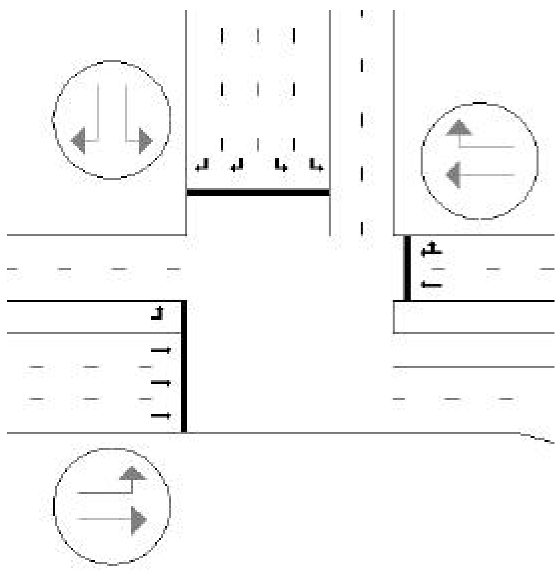

Figure 12 Geometry for Intersection #5

Running these intersections through the optimization and simulation models produced the following results. Table 4 shows the green times for each of the movements at each intersection.

Table 4 Pre- and Post-optimized Signal Timing (Five intersections Analysis)

Intersection	#1		#2		#3		#4		#5	
Movement	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
1	15	49	30	59	30	4	30	7	0	0
2	30	42	30	27	30	85	30	49	30	44
3	25	4	20	4	0	0	20	44	0	0
4	25	62	0	0	20	11	0	0	15	4
5	15	50	30	11	0	0	30	0	30	4
6	30	4	30	4	30	4	30	75	30	4
7	25	4	0	0	20	96	0	0	15	4
8	25	100	20	75	0	0	20	25	0	0

The five intersections, when simulated, were placed into the arterial system and the results were observed. There was no need to determine an offset since, as it was shown in the paper by the TTI [2], offset had no effect upon the intersections. Another reason is that the internal queue management takes the offset into account for each intersection. The results for these five intersections used the overall system averages to determine the effect that the new timing plan had on performance.

The average total time for the entire system in the pre-optimized condition was 91.9 seconds per vehicle, with 357 vehicles per hour. The post-optimized average total time was 65.3 seconds per vehicle, with 431 vehicles per hour. This is a decrease of 29 percent for the average total time and an increase of 21 percent for the average volume. The animation produced in *CORSIM* was viewed in order to visualize the performance. With the new timing plans, the traffic signals would only allow a certain number of vehicles on the internal links, and there was little to no spillback into the upstream intersections. This showed that some of the queue management techniques implemented in the formulation did indeed provide the appropriate control for the arterial.

Conclusions and Recommendations

This methodology has not yet been tested in the field, but the simulation modeling shows an increase in the operational performance of the arterial. By optimizing both the phasing sequence and phase timings, any oversaturated system should be able to have an improvement in the quality of service provided during the peak periods. While this does not altogether eliminate the oversaturation, it does provide a more effective means of processing the vehicles through the system. In the case study, the arterial performance showed definite improvement over the existing condition, and this was after the arterial was artificially made severely oversaturated.

The limits of the model, in terms of maximum degree of saturation, are yet unknown, and knowing these limits could provide a way for traffic engineers to optimize a system of oversaturated arterials even more efficiently. Further studies could be conducted, using several different scenarios to see what limitations the model might have. It is recommended that this methodology should be applied to a section of an actual arterial and the system performance observed. This could be done on a trial basis, for a brief period, to determine if the improvement shown in the simulation model will actually occur in the field. The case study intersections would be a good start for the testing of the methodology. Further testing could be performed on different sections of arterials to show how the response is for different input parameters to the model.

One of the constraints of this model is that it requires the use of the *HCS-3* program to get the initial capacity, and then it iterates this *HCS-3* process every time that the signal timing changes. Developing software that could optimize the signal timings and then recompute the capacity would correct this constraint of the model. The optimization would then also provide the best plan based upon the fluctuating capacity. This was not corrected in the linear programming, since it would have created non-linear constraints, which would require non-linear programming.

Since the area of oversaturation has not been studied extensively, there are many opportunities to further the research begun in this paper. This area is becoming increasingly important as the volume of traffic in cities grows, and could be beneficial to helping control some of the congestion that already exists in major cities.

In the process of optimizing the phase timings and cycle length, the capacity was also optimized, since the capacity is directly proportional to the green ratio. This presented some challenges, and further investigation is needed, to see the effects of the dynamic capacity upon the linear optimization. This process would be an iterative one, in which the linear program is run on the existing conditions in order to get the new phase times, and then the capacity is determined, and then the model is run again to see the effects upon the capacity.

PHASE II: CYCLE-BY-CYCLE ANALYSIS

INTRODUCTION

The development of effective traffic signal timing plans has long been a challenge for traffic engineers. Different control strategies can be developed based on the network geometric configuration and prevailing traffic characteristics, but the performance of a particular strategy cannot be fully predicted due to the stochastic nature of traffic flows. This can be clearly seen by examining the cycle-by-cycle variation in the traffic demand and vehicular arrival patterns for a given intersection. Normally, in actuated coordinated systems, signal parameters such as background cycle length and offsets are kept constant for a given time period. The signal parameters are typically determined based on the average traffic flow, which often causes spillback or starvation at the intersection when traffic arriving at the intersection exceeds the average flow. To overcome this limitation, these parameters need to be adjusted on a cycle-by-cycle basis according to the variation in the flow arriving at the intersection.

Currently, signal timing design strategies are developed and analyzed following procedures in the *Highway Capacity Manual (HCM)*, but these procedures do not provide methods for analyzing the signal systems on a cycle-by-cycle basis. Therefore, the objective of this study is to develop a methodology that selects a particular signal timing plan for an actuated coordinated intersection based on the fluctuations in the arrival rate for each cycle, while maintaining coordination with the adjacent intersections. In the proposed methodology, the congestion mechanism at each approach can be described quantitatively based on the volume arriving at different intervals of the cycle. The actuated signal control parameters are then determined sequentially on a cycle-by-cycle basis.

METHODOLOGY

The control delay is the measure of effectiveness commonly used to analyze the operation of signalized intersections. The control delay is defined as the delay created by both the red phase for the intersection, and the effect of the current traffic conditions. The control delay is defined in the HCM, 2000 update, by three different delay terms: uniform delay, incremental delay, and overflow delay. The control delay equation for a signalized intersection is given by the equation:

$$d = d1 * PF + d2 + d3$$

Where:

d	=	control delay (sec/veh);
d1	=	uniform delay (sec/veh);
PF	=	uniform delay progression adjustment factor;
d2	=	incremental delay (sec/veh); and
d3	=	initial queue delay (sec/veh).

The first term in the delay equation is the uniform delay multiplied by a progression factor that relates to how well the traffic is flowing along an arterial. This uniform delay estimation is based upon the Webster's delay formulation, and is widely accepted as an accurate representation of delay for an idealized arrival pattern. The second term in the delay equation is the random or incremental delay term. This term accounts for the non-uniform arrivals and individual cycle failures. The random delay is theoretically developed and is used to create a more realistic model of delay for the vehicles in the system. This term partially accounts for the delay created by the oversaturation of the intersection, but not for the residual queue at the start of the cycle. This delay term is sensitive to the degree of saturation in the system and the time period being analyzed.

The third term in the initial queue delay, which accounts for delay to all vehicles in the analysis period due to initial queue at the start of the analysis period.

While the *HCM2000* method of estimating delay can be used to analyze intersections operating under oversaturated conditions, it does not provide representation of the cycle-by cycle queue and delay dynamics and their relationship with the coordination and offsets between intersections in corridors. In order to evaluate the effectiveness of traffic signal coordination schemes, the signal-induced platoon delays must be analyzed.

The model presented in this research is based on the work of Rouphail [7]. It provides a more accurate modeling of queue and delay dynamics associated with platoon cycles over a sequence of cycles, with platoons of varying sizes, densities, and progression scenarios. The basis of this model is a cycle-by-cycle simulation of platoon arrivals, departures, and overflows. Within this model, vehicles are divided into the following categories:

- Residual arrivals from cycle i-1
- Arrivals in red in cycle i
- Arrivals in green in cycle i
- Beginning overflow queue in cycle i-1
- Residual arrivals in red from cycle i-1
- Platoon arrivals in cycle i

Once the number of vehicles in each category is identified, the task of computing delays is greatly simplified. This is because all vehicles in a category have a fixed arrival headway (h_a) and depart at either the same headway when no delays are incurred or at the saturation headway (h_d) when some delay takes place. Total delay is computed as the summation of the difference in departure and arrival times for all vehicles in a group, and then averaged over the number of vehicles in that group. The method, however, ignores variations in both arrival and departure headways and does not consider the effect of secondary flows on platoon delays. The following variables are used in the problem formulation. Entries labeled with asterisks are inputs to the model:

A_i = Scheduled demand in cycle i (= AR_i + AG_i)

AD_k = Generic variable designating the average overall delay for vehicles in category k.

AEG = Maximum platoon size (with headway ha) in which can proceed unimpeded until the end of each cycle, barring the presence of queues at the stop line

AG_i = Arrivals in green from platoon arriving in cycle i

AMAX* = Maximum platoon size, vehs

AMIN* = Minimum platoon size, vehs

AR_i = Arrivals in red from platoon arriving in cycle i

ARMX =	Maximum platoon size (with headway ha) which may arrive prior to the start of the green phase in each cycle
BOQ_i =	Begin overflow queue in cycle i (BOQ_1 defaults to zero)
C* =	Cycle length in seconds
c =	Movement capacity in veh/hr = 3600sg/C
d =	Average overall delay (including acceleration and deceleration delays) per vehicle in sec
DEM_i =	Net total vehicle demand in cycle i; from previous overflows + residuals from last cycle + scheduled arrivals in cycle i – residuals arriving in cycle i+1
DIS_i =	Net vehicle discharge in cycle i
EOQ_i =	End overflow queue in cycle i
g* =	Effective maximum green time in seconds
h_a* =	Average arrival headway platoon
h_d* =	Average departure headway at stop line
i =	Cycle designator, i=1,…,M
N_k =	Generic variable designating the number of vehicle in category k
OAG_i =	Portion of EOQ_i consisting of vehicle arrivals in green in cycle i
OAR_i =	Portion of EOQ_i consisting of vehicle arrivals in red in cycle i
OFS* =	Platoon Offset, measured from the start of the red phase
OO_i =	Portion of EOQ_i consisting of overflows from previous cycles
$ORAR_i$ =	Portion of EOQ_i consisting of residual arrivals from cycle i-1
QRED=	Queue length at the end of the red phase (veh)
r =	Effective red time in secs (= C-g)
RAR_i =	Residual arrivals in red in cycle i+1; belong to platoon (leader) arriving in cycle i
s =	Saturation flow rate in veh/s
T =	Flow period, in hrs
t_i =	Time first residual vehicle from cycle i-1 arrives in cycle i red phase
TD_k =	Generic variable designating the total delay for vehicles in category k
X =	Degree of saturation

X_o = Degree of saturation below which overflow delay is considered negligible

Arrivals, Discharges, and Queues:

$d = d1 * PF + d2 + d3$

AEG = Integer [(C – OFS) / h_a +1]

ARMX= Max {0, Integer [(r-OFS) / h_a +1]}

A_i = ROUND {AMIN + RND*(AMAX – AMIN)}

RAR_i = Max {0, A_i-AEG}

AG_i = Max {0, A_i – RAR_i}, when OFS >= r

= Max {0, A_i-RAR_i – ARMX}, when OFS < r

AR_i = A_i – RAR_i – AG_i

BOQ_i = EOQ_{i-1}, i=2,….,M (BOQ_1 defaults to zero)

DEM_i = BOQ_i + RAR_{i-1} + A_i – RAR_{iI}

DIS_i = Min {Integer (g/h_d), DEM_i}

EOQ_i = Max {0, DEM_i – DIS_i}

$QRED_i$ = BOQ_i + AR_i

OO_i = Max {0, BOQ_i – DIS_i}

$ORAR_i$= 0 if EOQ_i <= A_i – RAR_i = EOQ_i – A_i + RAR_i otherwise

OAG_i = Min {EOQ_i, AG_i}

OAR_i = Min {AR_i, Max [0, EOQ_i – AG_i]}

OA_i = OAG_i + OAR_i

t_i = OFS + h_a*AEG – C, if RAR_{i-1} > 0

Delay To Vehicles Discharging in Cycle i:

From beginning overflow queue BOQ $_i$

$AD_i°$ = Average delay per vehicle in group

$AD_i°$ = Average delay for <u>all</u> overflow vehicles in cycle (i-1); it is added to the current cycle delay on the basis of the number of vehicles in the group $N_i°$

$N_i°$ = Number of vehicles in group

TD_i° = Total delay for group

$N_i^\circ = BOQ_i - OO_i$

$TD_i^\circ = N_i^\circ \{r-h_d\} + 0.50\, N_i^\circ(N_i^\circ+1)h_d + AD_{i-1}^\circ\, N_i^\circ$

$AD_i^\circ = TD_i^\circ / N_i^\circ$

From residual arrivals from cycle i-1

AD_i^{RAR} = Average delay per vehicle in group

N_i^{RAR} = Number of vehicles in group

TD_i^{RAR} = Total delay for group

$N_i^{RAR} = RAR_{i-1} - ORAR_i$

$TD_i^{RAR} = N_i^{RAR} \{r + BOQ_i\, h_d + h_a - t_i - h_d \} - 0.50\, N_i^{RAR} (N_i^{RAR} +1) (h_a - h_d)$

$AD_i^{RAR} = TD_i^{RAR} / N_i^{RAR}$

From arrivals in red in cycle i

AD_i^{AR} = Average delay per vehicle in group

N_i^{AR} = Number of vehicles in group

TD_i^{AR} = Total delay for group.

$N_i^{AR} = AR_i - OAR_i$

$T_i^{AR} = N_i^{AR} \{r + h_d (BOQ_i+RAR_{i-1}) + h_a - OFS - h_d\} - 0.50 N_i^{AR}(N_i^{AR} +1) (h_a - h_d)$

$AD_i^{AR} = TD_i^{AR} / N^{AR}$

From arrivals in green in cycle i

AD_i^{AG} = Average delay per vehicle in group

N_i^{AG} = Number of vehicles in group

TD_i^{AG} = Total delay for group

If OEQ_i > zero, then: $N_i^{AG} = AG_i - OAG_i$

$TD_i^{AG} = N_i^{AG} \{r + h_d (BOQ_i + RAR_{i-1} + AR_i) + h_a (1 - AR_i) - OFS - h_d\} - 0.50\, N_i^{AG} (N_i^{AG} + 1) (h_a - h_d)$

If OEQ_i = zero, then:

I_i = Integer $\{[r + h_d (BOQ_i + RAR_{i-1}) + h_a - OFS - h_d] / [h_a - h_d] - AR_i$

$N_i^{AG} = \text{Min}(I_i, AG_i)$

$TD_i^{AG} = N_i^{AG} \{r + h_d (BOQ_i + RAR_{i-1} + AR_i + h_a (1 - AR_i) - OFS - h_d\} - 0.50\, N_i^{AG} (N_i^{AG} + (h_a - h_d)$

$TD_i^{AG} = N_i^{AG} \{r + h_d (BOQ_i + RAR_{i-1} + AR_i) + h_a (1 - AR_i) - OFS - h_d\} - 0.50\, N_i^{AG} (N_i^{AG} + 1)(h_a - h_d)$

$AD_i^{AG} = TD_i^{AG} / N_i^{AG}$

Delays to Vehicles Overflowing to Cycle i+1

From beginning overflow queue in cycle i

AD_i^O = Average delay per vehicle in group

N_i^O = Number of vehicles in group

TD_i^O = Total delay for group

$N_i^O = OO_i$

$TD_i^O = CN_i^O * (C + ADi_{-1})$

$AD_i^O = TD_i^O / N_i^O$

From residual arrivals in red from cycle i-1

AD_i^{RAR} = Average delay per vehicle in group

N_i^{RAR} = Number of vehicles in group

TD_i^{RAR} = Total delay for group

$N_i^{RAR} = ORAR_I$

$TD_i^{RAR} = N_i^{RAR} \{C - t_i - h_a [RAR_{i-1} - 1 - ORAR_i]\} - 0.50 * h_a N_i^{RAR} \{N_i^{RAR} + 1\} + N_i^{RAR} * AD_{i-1}$

$AD_i^{RAR} = TD_i^{RAR} / N_i^{RAR}$

From platoon arrivals in cycle i

AD_i^A = Average delay per vehicle in group

N_i^A = Number of vehicles in group

TD_i^A = Total delay for group

$N_i^A = OAR_i + OAG_i$

$TD_i^A = N_i^A \{C - OFS - h_a [AR_i + AG_i - OA_i - 1] \} - 0.50 * h_a N_i^A \{ N_i^A + 1\} + N_i^A * AD_{i-1}$

$AD_i^A = TD_i^A / N_i^A$

SIGNAL CONTROL ALGORITHM

The signal control algorithm used in this research was designed to prevent spillback at different links by controlling the queue length, and to minimize the total delay at the intersection. The computation was performed on a cycle-by cycle basis using the delay estimates for vehicles arriving at different intervals throughout the cycle. The intersection signal control parameters are sequentially controlled, so the queue length at the end of the red phase is less than the storage capacity of the arterial link ($QRED_{i, K} <= STORAGE_K$). The signal control parameters are determined as follow:

Step 1: Signal control parameters (offset, maximum green, and cycle length) are set based on the traffic conditions for cycle i-1

Step 2: Detector data are used along with the signal control parameters to determine the number of vehicles arriving at each of the following intervals during the cycle i:

- Residual arrivals from cycle i-1
- Arrivals in red in cycle i
- Arrivals in green in cycle i
- Beginning overflow queue in cycle i-1
- Residual arrivals in red from cycle i-1
- Platoon arrivals in cycle i

Step 3: The queue length at the end of the red phase is compared with the arterial storage capacity. If the computed queue length satisfies the conditions ($QRED_{i, K} <= STORAGE_K$), the signal control parameters are implemented, otherwise the analysis will continue to the next step.

Step 4: The duration of the red phase for the arterial is corrected to satisfy the condition ($QRED_{i,K} <= STORAGE_K$)

Step 5: New green times for the arterial are assigned based on the total demand during the cycle (DEM_i). The cycle length for the arterial is adjusted accordingly, to reflect the new control parameters.

Step 6: An optimal relative offset for the arterial is then assigned based on the length of the queue stored ($QRED_i$)

Step 7: Steps 2 through 6 are repeated using the new signal control parameters.

Step 8: The cycle-by-cycle delay is estimated for vehicles arriving in different intervals. The total delay for the arterial is also estimated.

Step 9: Maximum green time for other approaches and phase sequence are then optimized using the linear optimization procedure introduced in the first phase of this research.

EVALUATING THE MODEL THROUGH SIMULATION

The use of simulation programs to evaluate traffic operations strategies is growing in popularity worldwide. Traffic simulation programs offer an advantage over traditional analysis tools by providing both an analysis of the entire network and a visual simulation of the results. All of the required traffic characteristics, such as traffic volumes, turn movements, traffic regulations, signal timing and traffic geometry, are entered into the simulation software as inputs and the software controls the model based on these inputs.

The TSIS/CORSIM program that was developed by FHWA is one of the most commonly-used simulation programs. Though the results of CORSIM simulations are extremely useful, the program employs actuated traffic control technology from the 1980s, and does not allow the user to specify the advanced control algorithms that have been developed in recent years. Using a process called real-time hardware-in-the-loop simulation, engineers can test these new algorithms by linking a late-model, actual controller to the CORSIM simulation model. This direct communication between a controller and the CORSIM software is made possible by a device called the Controller Interface Device (CID), developed by NIATT at the University of Idaho.

In this study, the simulation model developed in this research was simulated using CORSIM alone and also using CORSIM with the CID in hardware-in-the-loop simulation. A comparison of the results of the two sets of simulation will help validate the hardware-in-the-loop simulation model used in the analysis. The following paragraphs describe these two types of simulations.

CORSIM SIMULATION

CORSIM is a microscopic integrated simulation program. It applies time-step simulation, with one step equaling one second, to describe traffic operations. In the program, each vehicle is a distinct object that is moved every second. Each variable control device (such as a traffic signal) and each event are updated every second. CORSIM and other traffic simulation programs are based on stochastic algorithms that describe driver behavior and traffic operations, and rely on a random number seed to generate vehicles. The measures of effectiveness (MOEs) that are obtained from a simulation such as total time, delay time, queue time, and vehicle discharged are the result of a specific set of random number seeds. Because of the stochastic nature of the model, it is necessary to perform several simulation runs, and report the average result as well as the amount of variation in the result.

CORSIM’s output includes several different measures of effectiveness. Of these, the total time, delay time, queue time in sec per vehicle, and the discharge per approach were considered in this research, for the purpose of comparison. CORSIM defines total time, delay time, and queue time as follows:

- Total Time per vehicle (sec per veh) - The average travel time on a link for each vehicle, calculated by taking the total travel time and dividing it by the number of vehicle trips.
- Delay Time per vehicle (sec per veh) - The average delay on a link for each vehicle, calculated by taking the delay time in veh-min and dividing it by the number of vehicle trips.

- Queue Delay per vehicle (sec per veh) - Delay calculated by taking vehicles having acceleration rates less than 2 feet per second^2 and speed less than 9 feet per second. If a vehicle's speed is less than 3 feet per second, it will be included every second. Otherwise it will be included every two seconds.

HARDWARE-IN-THE-LOOP SIMULATION

In hardware-in-the-loop traffic simulation, the traffic controller component of the simulation (the internal controller emulation logic) is replaced with a real traffic signal controller. The model runs in real time, i.e. one second of simulation takes one second of actual time, since controller hardware usually performs input and output in real time.

To achieve this external simulation control, the simulation software and the controller hardware must be able to communicate. The CID and its accompanying software are the tools that allow this communication.

In a case of testing a model with an actuated controller, a hardware-in-the-loop simulation consists of the following steps:

1. CORSIM generates detector actuations by modeling simulated vehicles crossing simulated detectors.
2. The simulated detector actuations are sent to the actual controller hardware.
3. The controller reacts to them as it would react to real detector actuations, by updating phase indications according to the phasing and timing plan programmed in the controller.
4. The phase indications are subsequently read back from the controller hardware to CORSIM, and assigned to the simulated traffic signals.
5. The simulated vehicles then react to the simulated traffic signals by stopping or departing as appropriate.

This hardware-in-the-loop function provides two major benefits to traffic engineers:

- It allows engineers to evaluate and fine-tune the signal timing parameters in an actual, late-model NEMA or 170 traffic controller.
- It allows the testing to take place right in the office, rather than out in the field, where traffic could possibly be disrupted.

It is also believed that hardware-in-the-loop simulation can be used to reliably evaluate the performance gains associated with special features of the various traffic controllers, but that possibility has not been examined in this research.

SIMULATION MODEL DEVELOPMENT

Three closely spaced intersections in the city of Moscow were selected for this research: Third & Jackson Street, Third & Main Street and Third & Washington Street. A link-node model for the network is presented in Figure 13; the network geometric configuration is presented in Figure 14.

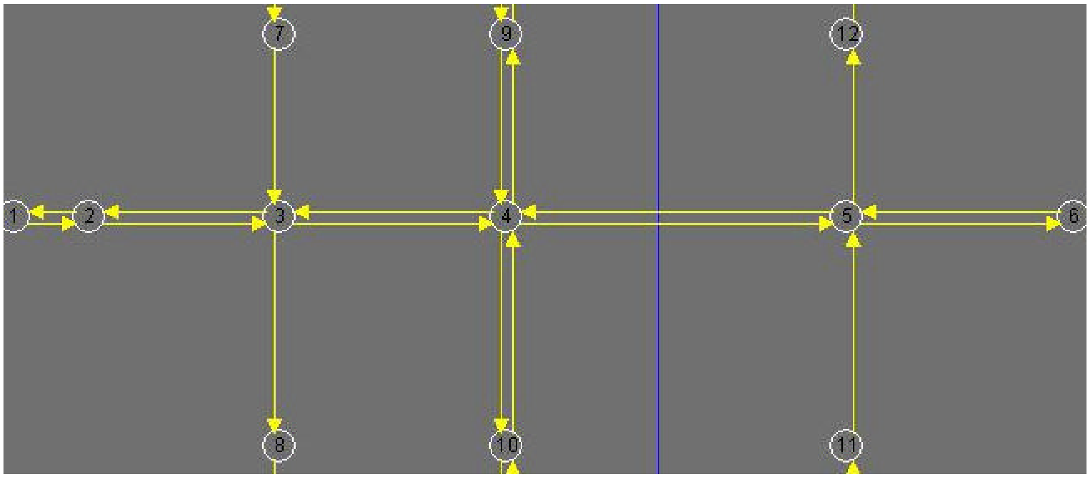

Figure 13 Link-Node diagram for the network

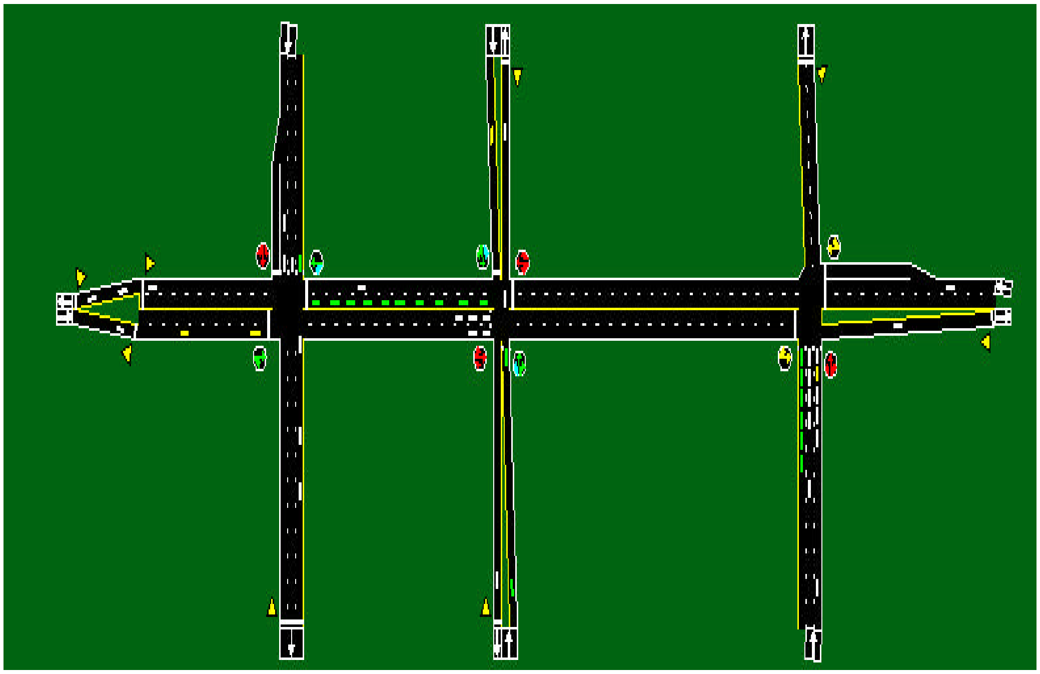

Figure 14 Network Configuration

Testing the Model with CORSIM

The simulation model developed for the arterial was run and the simulation animation was used to verify the inputs and confirm vehicle movement and signal operations. A TSIS multi-run same case script was used to perform multiple CORSIM simulations of the network using different random number seeds to generate vehicles. A normal headway distribution was used for all of these runs. The relevant measures of effectiveness (total time, delay time, queue length and vehicle discharge) were recorded and compared against field data to validate the model. Results of the validation are presented in Table 5.

Table 5 Simulation Model Validation Results

Link	Vehicle output			Average Queue Length		
	CORSIM	Field Data	percent Difference	CORSIM	Field Data	percent Difference
3-4	543	558	-2.69	4.6	5.2	-11.54
5-4	571	592	-3.55	6.3	5.8	8.62
9-4	106	123	-13.82	5.2	4.9	6.12
10-4	233	218	6.88	4.3	4.6	-6.52

Testing the Model with Hardware-in-the-Loop Simulation

The next step was to run the hardware-in-the-loop simulation model using CORSIM with the CID, to test the model performance under actual NEMA TS1 controller in a real-time simulation environment. Figure 15 presents the setting for the hardware-in-the-loop simulation model used in the analysis.

The CORSIM input file describes the traffic flow, geometry, and control parameters of the traffic system being modeled. One of the CID software applications, *the CID Configuration Tool*, modifies this input file so that it can be used with the CID. The *CID Configuration Tool* also creates a “Configuration File,” which tells CORSIM which intersections will be controlled by the actual controller and which nodes will be connected to which CID.

Once the necessary support files had been generated, the hardware-in-the-loop simulation was conducted for ten one-hour runs. The random number seeds were again repeated for each run within each volume. All of the parameters, including the arrival pattern, were the same for both the CORSIM and CID simulation runs.

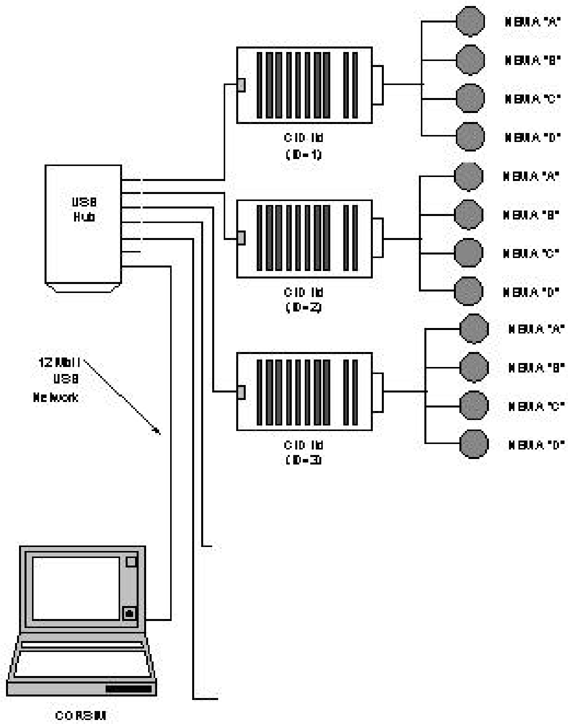

Figure 15 Hardware-in-the-Loop Simulation Model Setting

COMPARISON BETWEEN CORSIM AND HARDWARE-IN-THE-LOOP SIMULATION MODELS

The output from the CORSIM simulation and the hardware-in-the-loop simulation runs were analyzed. The total time, delay time, queue time, and number of vehicles discharged from each link for the two simulation models are presented in Tables 6 through 9. In general, while the results of the two models were comparable, there is a difference between the measures obtained from the two models. These differences are attributed to the difference between the operation of the actual controller used in the hardware-in-the-loop simulation and the generic controller emulator used in CORSIM. There was no direct relationship or apparent trend between the results obtained from the two simulation models.

Table 6 Total Travel Time (CORSIM versus Hardware-in-the-loop simulation)

Run	Link (3,4)			Link (5,4)			Link (9,4)			Link (10,4)		
	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff
1.0	24.3	26.3	-7.5	22.3	22.2	0.6	23.2	22.5	3.1	38.1	36.7	3.8
2.0	27.1	26.8	1.1	23.7	24.6	-3.8	20.6	19.8	4.0	36.9	38.6	-4.4
3.0	24.3	28.1	-13.5	21.7	22.2	-2.2	23.3	23.9	-2.5	37.6	41.1	-8.5
4.0	28.1	29.9	-6.0	25.1	23.9	4.9	23.0	21.0	9.5	40.0	39.4	1.5
5.0	24.3	22.4	8.6	22.7	20.3	11.8	21.3	23.5	-9.2	36.6	35.0	4.6
6.0	25.6	23.2	10.4	21.7	20.3	6.8	21.2	21.3	-0.5	37.7	37.0	1.9
7.0	26.2	28.0	-6.5	23.2	22.7	2.3	24.9	22.6	10.2	33.7	35.8	-5.9
8.0	23.7	21.9	8.0	20.5	21.3	-3.9	22.6	20.0	13.0	32.1	34.0	-5.6
9.0	24.2	25.1	-3.5	23.9	24.2	-1.2	23.2	25.0	-7.1	38.7	41.4	-6.7
10.0	22.7	21.9	3.6	22.1	24.4	-9.3	23.0	22.3	3.1	33.2	36.3	-8.5
Mean =	25.1	25.4	-1.2	22.7	22.6	0.4	22.6	22.2	2.0	36.5	37.5	-2.9

Table 7 Average Delay (CORSIM versus Hardware-in-the-loop simulation)

Run	Link (3,4)			Link (5,4)			Link (9,4)			Link (10,4)		
	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff
1	16.3	17.1	-4.7	16.3	14.7	10.9	18.6	18	3.3	16.8	13.9	20.9
2	16.2	16.9	-4.1	14.9	14.4	3.5	17.8	15.2	17.1	13	12.9	0.8
3	17.5	17.5	0.0	12.5	13.5	-7.4	17.7	19.3	-8.3	13.9	15.5	-10.3
4	14.9	16.6	-10.2	14.9	15	-0.7	18.5	16.4	12.8	14.3	13.7	4.4
5	16.2	15.6	3.8	14.6	16.1	-9.3	15.1	17	-11.2	18.3	19.2	-4.7
6	14.6	16.4	-11.0	15.4	16.2	-4.9	16.7	16.8	-0.6	15.2	13.3	14.3
7	15.2	17.1	-11.1	15.4	17.9	-14.0	19.3	18	7.2	16.7	16.2	3.1
8	16.3	15.1	7.9	14.6	16.7	-12.6	19.5	15.4	26.6	14.4	18.4	-21.7
9	16.2	18.3	-11.5	17.2	18.2	-5.5	20.5	20.4	0.5	16.5	14.7	12.2
10	14.9	15.1	-1.3	15.9	16.3	-2.5	18.7	17.8	5.1	15.6	13.5	15.6
Mean =	15.8	16.6	-4.5	15.2	15.9	-4.6	18.2	17.4	4.6	15.5	15.1	2.2

Table 8 Queue Time (CORSIM versus Hardware-in-the-loop simulation)

Run	Link (3,4)			Link (5,4)			Link (9,4)			Link (10,4)		
	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff
1	5.9	6.3	-6.3	8.4	8.6	-2.3	7.2	6.9	4.3	13.9	10.9	27.5
2	5.8	6.1	-4.9	8.1	10.7	-24.3	6.6	8.1	-18.5	10.4	8.4	23.8
3	5.3	6.3	-15.9	9.2	8.8	4.5	8.3	9.6	-13.5	11.9	10.4	14.4
4	7.3	9.2	-20.7	10.9	10.6	2.8	8.2	8.4	-2.4	10.7	9.2	16.3
5	6.6	6.1	8.2	8.7	7.9	10.1	7.6	7.1	7.0	11.6	10.1	14.9
6	7.8	7.9	-1.3	11.4	12.3	-7.3	7.7	8.1	-4.9	10.9	8.9	22.5
7	7.3	7.3	0.0	8.7	10.9	-20.2	7.8	6.2	25.8	12	10.2	17.6
8	5.8	7.1	-18.3	9.6	9.2	4.3	7.7	8.7	-11.5	11.7	12.3	-4.9
9	5.9	4.9	20.4	10.1	10.9	-7.3	8.9	8.3	7.2	12.5	10.6	17.9
10	5.7	6	-5.0	8.2	9.6	-14.6	7.2	6.8	5.9	12.8	10.9	17.4
Mean =	6.3	6.7	-5.7	9.3	10.0	-6.2	7.7	7.8	-1.3	11.8	10.2	16.2

Table 9 Volume Discharge by Link (CORSIM versus Hardware-in-the-loop simulation)

Run	Link (3,4)			Link (5,4)			Link (9,4)			Link (10,4)		
	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff	CID	CORSIM	percent diff
1	544	553	-1.654	567	572	-0.882	106	106	0	233	233	0
2	547	551	-0.731	571	576	-0.876	106	105	0.943	232	230	0.862
3	543	542	0.184	569	574	-0.879	105	108	-2.857	233	233	0
4	544	542	0.368	568	565	0.528	106	106	0	233	233	0
5	546	548	-0.366	573	566	1.222	106	106	0	231	231	0
6	594	549	7.576	566	571	-0.883	108	105	2.778	235	231	1.702
7	547	550	-0.548	572	573	-0.175	106	106	0	233	232	0.429
8	545	551	-1.101	566	571	-0.883	106	105	0.943	233	231	0.858
9	552	552	0	574	573	0.174	105	105	0	232	232	0
10	546	548	-0.366	571	573	-0.35	106	106	0	233	232	0.429
Mean =	550.8	548.6	0.399	569.7	571.4	-0.298	106	105.8	0.189	232.8	231.8	0.43

EVALUATION OF OPTIMIZATION STRATEGY

The signal control parameters obtained through the cycle-by-cycle control algorithm were tested using the validated hardware-in-the-loop simulation model. To examine the delay reduction potential of the proposed cycle-by-cycle control method it was compared against the base case, an optimized actuated coordinated signal system.

The first part of the analysis was to develop an optimized actuated coordinated timing plan for the corridor. Initial plans were developed using macroscopic optimization tools, *TRANSYT7-F* and *SYNCRHO 4.0*. The resultant timing plans were further analyzed and the performance of the corridor under these plans was tested using the hardware-in-the-loop simulation model. The results of the simulation model helped identify the optimal timing plans for the corridor. The optimal background cycle length for the corridor was 80 seconds. Twenty-five multiple one-hour simulation runs using different random seed numbers were then performed, to determine the measures of effectiveness for the corridor under the optimal actuated coordinated plan (the base case).

The next step in the analysis was to run the simulation model under the cycle-by-cycle control method. The same number of multiple runs using the same random seed numbers was performed and the corridor measures of effectiveness were obtained and compared against those for the base case.

In general, the proposed cycle-by-cycle control method was effective in eliminating the spillback and overflow problems. While spillback occurred in an average of 8.1 cycles during the one-hour simulation period under the base case control, no spillback occurred under the cycle-by-cycle control. Similarly, overflow was experienced in 11.2 percent of the cycles under the base case control, while under the cycle-by-cycle control only 3.8 percent of the cycles had overflow vehicles. Under the cycle-by-cycle control, the average delay for the major traffic in the corridor (Eastbound/Westbound) was reduced by an average of 8.23 percent. The average delay for the minor traffic was increased by an average of 11.1 percent. However, the overall average delay for the intersection was reduced by 6.1 percent. The pre-optimized and post-optimized average queue length, average delay, average queue time, and link ditching are presented in Tables 10 and 11.

Table 10 Pre-optimized and Post-optimized Average Queue Length and Average Delay

Link	Average Queue Length (veh)			Average Delay (sec/veh)		
	Pre-Optimized	Post-Optimized	percent Difference	Pre-Optimized	Post-Optimized	percent Difference
3-4	4.6	3.42	-25.65	15.8	14.11	-10.70
5-4	6.3	5.1	-19.05	15.2	14.16	-6.84
9-4	5.2	6.3	21.15	15.1	16.21	7.35
10-4	4.3	5.2	20.93	17.4	19.03	9.37

Table 11 Pre-optimized and Post-optimized Average Queue Time and Link Discharge

Link	Average Queue Time (sec/veh)			Vehicle Discharged (vph)		
	Pre-Optimized	Post-Optimized	percent Difference	Pre-Optimized	Post-Optimized	percent Difference
3-4	6.3	5.59	-11.27	550	593	7.82
5-4	9.3	8.44	-9.25	569	609	7.03
9-4	7.7	8.69	12.86	165	150	-9.09
10-4	11.8	12.73	7.88	231	199	-13.85

CONCLUSIONS AND RECOMMENDATIONS

This part of the report presents a methodology to prevent spillbacks and overflows of congested corridors using a cycle-by-cycle control algorithm. The algorithm is based on the cycle-by-cycle estimation of delay and queue length for vehicles arriving during different time intervals during the cycle. The proposed control method has not yet been tested in the field; rather, it was tested under a real NEMA TS2 controller environment using a hardware-in-the-loop simulation model. Results from the simulation modeling shows an increase in the operational performance of the arterial. By preventing spillback and overflow on the main arterial, the proposed method showed improvement in the quality of service provided during the peak periods. The proposed method provided an effective way of processing the vehicles through the system. In the case study examined in this research, the arterial performance showed improvement over coordinated actuated signal systems. The average delay for the arterial was reduced by 8.23 percent. The average delay for the minor traffic was increased by an average of 11.1 percent, but the overall average delay for the intersection was reduced by 6.1 percent.

The limits of the model and its applicability in the field are yet unknown, and knowing these limits could provide a way for traffic engineers to optimize a system of oversaturated arterials even more efficiently. Further studies could be conducted, using different traffic volumes and network configuration, to see limitations that the model might have or to more accurately assess its delay reduction potential.

REFERENCES

1. Kuzbari, Ray. "Linear Programming Optimization of Actuated Signal Timings Under Oversaturated Conditions." ITE Journal on the web (May 1998).

2. Kim, Youngchan, and Messer, Carroll, J. "Traffic Signal Timing Models for Oversaturated Signalized Interchanges." Research Report 1148-2. Texas Transportation Institute. Austin, Texas. (1992).

3. Abu-Lebdah, G., and Benekohal, R. "Development of Traffic Control and Queue Management Procedures for Oversaturated Arterials." *Transportation Research Record* 1603. Transportation Research Board. National Research Council. Washington, D. C. (1997).

4. Rouphail, Nagui, and Khatib, Zaher. "Arterial Signal Optimization Considering Left-Turn Control." *Transportation Research Record* 1456. Transportation Research Board. National Research Council. Washington, D. C. (1994).

5. Schrage, Linus. *Optimization Modeling with LINDO*, 5th Edition. Duxbury Press. (1997).

6. Calvert, James, and Voxman, William "Linear Programming". Harcourt Brace Jovanovich, Inc. (1989).

7. Rouphail, Nagui "A Cycle-Based Method for Estimating Platoon Delay". Working paper, Urban Transportation Center, The University of Illinois at Chicago, Chicago, IL. (1990)

Software Used

1.*LINDO 6.0* for Windows, LINDO Systems Inc. 1996.

2.*TSIS 4.32. (CORSIM)*, ITT Industries Systems Division. October 1999.

3.*HCS-3.2*, McTrans Center, University of Florida. 1998.

4.*Synchro 4.0*, Trafficware Corporation. 2000.