ABSTRACT

A CONCEPTUAL FRAMEWORK FOR USING FEEDBACK CONTROL WITHIN ADAPTIVE TRAFFIC CONTROL SYSTEMS

by
Renu Chhonkar

Existing adaptive traffic control strategies lack an effective evaluation procedure to check the performance of the control plan after implementation. In the absence of an effective evaluation procedure, errors introduced in the system such as inaccurate estimates of arrival flows, are carried forward in time and reduce the efficiency of the traffic flow algorithms as they assess prevalent traffic conditions. It is evident that the feed-forward nature of these systems cannot accurately update the estimated quantities, especially during oversaturated conditions.

This research is an attempt to develop a conceptual framework for the application of feedback control within the basic operation of existing adaptive traffic control systems to enhance their performance. The framework is applied to three existing adaptive traffic control strategies (SCOOT, SCATS, and OPAC) to enable better demand estimations and queue management during oversaturated condition. A numerical example is provided to test the performance of an arterial in a feedback environment. The example involves the design and simulation test of Proportional (P) and Proportional-Integral (PI) controllers and their adaptability to adequately control the arterial. A sensitivity analysis is further performed to justify the use of a feedback control system on arterials and to choose the type of controller best suited under given demand conditions. The simulation results indicated that for the studied arterial, the PI controller can handle demand estimation and

queuing better than P controllers. It was determined that a well designed feedback control system with a PI controller can effectively overcome some of the deficiencies of existing adaptive traffic control systems.

A CONCEPTUAL FRAMEWORK FOR USING FEEDBACK CONTROL WITHIN ADAPTIVE TRAFFIC CONTROL SYSTEMS

by
Renu Chhonkar

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
In Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Transportation

Interdisciplinary Program in Transportation

January 2004

BIOGRAPHICAL SKETCH

Author: Renu Chhonkar

Degree: Doctor of Philosophy

Date: January 2004

Undergraduate and Graduate Education:

Doctor of Philosophy in Transportation, 2004
New Jersey Institute of Technology, Newark, NJ

Master of Engineering in Civil Engineering, 2000
Maharaja Sayajirao University of Baroda, Baroda.

Bachelor of Engineering in Civil Engineering, 1998
Maharaja Sayajirao University of Baroda, Baroda.

Major: Transportation

Presentations and Publications:

Chien, I.S., Spasovic, L.N., Elefsiniotis, S.S., Chhonkar, R.S. *Evaluation of Feeder Bus Systems with Probabilistic Time-Varying Demands and Non-Additive Time Costs.* Transportation Research Record, No. 1760, 2001.

This dissertation is dedicated

to my Mother

ACKNOWLEDGEMENT

I wish to express sincere appreciation to my advisor Dr. Janice Daniel, for her moral support, encouragement and patience during this research. I am indebted to my advisor for reviewing my dissertation and offering her valuable comments and suggestions. It was a pleasure working with her.

I would also like to thank the members of my dissertation committee, Drs. Athanassios Bladikas, Lazar N. Spasovic, Steven Chien and Rachel Liu for their helpful corrections and productive comments. I am extremely fortunate to have them in the dissertation committee.

Finally, I shall be eternally grateful to my mother Gyan, father Suraj Bhan, and brother Anup for their love and inspiration through all these years. Thanks to my husband Sumit for his cooperation during this research. I thank my family for placing their faith in me and making this work possible.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

CHAPTER 1

INTRODUCTION

1.1 Background

Urban vehicular traffic is an expression of human behavior and as such, it is highly variable in space and time *(Gartner, 1985)*. Therefore, a high degree of adaptiveness is required in traffic management systems for handling these variable traffic conditions with reasonable efficiency. The concept of adaptive traffic control was introduced as a result of the inefficiency of fixed time signals to handle such varying demand conditions. A number of strategies for adaptive control have evolved and been tested since the 1970s with the advent of computer systems. Some of the well known strategies that have been brought into implementation are Split Cycle Offset Optimization Technique (SCOOT); Sydney Coordinated Adaptive Traffic System (SCATS); and OPAC (Optimization Policies for Adaptive Control).

Most existing adaptive control strategies are of a feed-forward nature. This means that the strategies follow a sequence of activities starting with data collection from online or offline sources and demand estimation, a decision making process to calculate the signal timing plan for the estimated demand, implementation of the timing plan, a verification process to check the correct implementation of the plan, and finally an evaluation of the strategy. The evaluation of the system is done by calculating various performance indices that describe the operation of the system. In the entire process, one activity leads to another in a sequential manner. The activity that occurs later has no effect on the activity occurring earlier in the sequence of events.

1

Although existing demand responsive strategies and techniques based on this operational sequence have been designed to address the problem of oversaturated traffic conditions and queue management through highly elaborate computerized systems, a number of shortcomings have been identified with the efficiency of these systems. Since most of these strategies are feed-forward in their operation, there is no corrective mechanism to adjust to the dynamic traffic flow patterns in different links of the network and estimate demand with reasonable accuracy.

One strategy that has a potential to enhance the function of existing adaptive control systems and make them less prone to errors is by operating these traffic control systems in a feedback environment. Feedback control systems have been applied to a number of engineering fields. These systems have been used to regulate or track target values in electrical, chemical, mechanical engineering, and operations research systems, to name a few. However, the use of control systems in the field of transportation has been relatively sparse and few attempts have been made to use control theory to regulate the performance of transportation facilities such as arterials and intersections.

This research is an attempt to develop a conceptual framework for the application of feedback control within the basic operation of existing adaptive traffic control systems to enhance their performance and efficiency. Using the feedback property available to control systems, existing adaptive systems can be modified so that the performance of the signal timing plan can be determined before it is implemented.

1.2 Problem Statement

Existing adaptive control systems are equipped with sophisticated algorithms to handle a variety of traffic conditions and avoid congestion. Despite the existence of these algorithms, there are conditions or combinations of conditions during which the adaptive control systems produce suboptimal results. It has been observed that the current adaptive control systems lack an effective evaluation procedure to check the performance of the control plan after implementation. In the absence of an effective evaluation procedure, errors introduced in the system such as inaccurate estimates of arrival flows, are carried forward in time and reduce the efficiency of the traffic flow algorithms as they assess prevalent traffic conditions. It is evident that the feed-forward nature of these systems can not update the estimated quantities in case of unpredictable conditions occurring either at upstream or downstream locations.

Oversaturation on arterials is one traffic condition which produces suboptimal signal control plans under most existing adaptive control systems. As soon as oversaturation begins, errors are introduced in the system in the form of queues and spillover. In such cases, the predicted values of demand by the traffic flow algorithm become invalid, and an update in the demand is required because the actual demand includes the predicted demand as well as residual queues. However, due to the lack of an efficient evaluation procedure to determine whether the plan is suboptimal, and, the lack of a feedback mechanism to adjust the control plan in subsequent time periods, the adaptive control systems fail to make the necessary changes to the control plan to serve the changing demand efficiently. Therefore, the errors introduced at one time step propagate, some times leading to a complete failure of the system.

In this research, feedback control is introduced as an approach to enhance the performance of existing adaptive control strategies by overcoming some of the limitations of a feed-forward system. Depending on the type of adaptive control algorithm, a feedback loop can be introduced in the basic operation of an adaptive control system to regulate the overall function of the system in an efficient manner. Depending on when the feedback loop is introduced in the adaptive control algorithm, specific objectives can be achieved, such as a more accurate estimation of demand through the inclusion of existing residual queues from oversaturated conditions. Thus, the introduction of a feedback loop can actively contribute in the timely update of the system parameters depending on the level of errors introduced in the systems.

1.3 Research Objectives

Incorporating a feedback control in adaptive traffic control systems can enable researchers to solve some of the problems associated with existing techniques of adaptive control at signalized intersections. This research is an attempt to use feedback control theory to develop an approach to improve the performance of adaptive traffic control systems within an arterial.

The objectives of this research are as follows:

- To develop a framework for enhancing the performance of adaptive control systems.

- To apply the framework to three existing adaptive traffic control strategies: SCOOT, SCATS, and OPAC.

- To demonstrate the efficiency of feedback control systems in improving the performance of signalized arterials with the help of a numerical example; and

- To test the performance and sensitivity of the proposed feedback control system under a Proportional (P) regulator and Proportional-Integral (PI) regulator.

1.4 Methodological Approach

The methodological approach followed in this research starts with a detailed study of the function of three adaptive traffic control strategies, namely, SCOOT, SCATS and OPAC. The study identifies conditions which lead to suboptimal performance of each of these systems. The research further proposes a conceptual framework for using feedback control into the operation of these adaptive traffic control strategies and individual frameworks. The enhancements expected by the implementation, such as reduction in queue lengths, are also studied.

It is out of the scope of this research to test the performance improvement of SCOOT, SCATS and OPAC due to the incorporation of a feedback control loop as this would require either a field implementation of these control strategies or the development of a simulation model capable of implementing adaptive control strategies. Instead, the performance of a roadway network in a feedback environment is tested and results are analyzed with the help of MATLAB (MATrix LABoratory). This study tests the performance of a hypothetical roadway arterial in a feedback environment using P and PI feedback controllers. The P and PI controllers, depending on their individual characteristics, attempt to reduce the errors of the feedback control system and make the system track a desired output. The performance of the study arterial is analyzed under the influence of feedback controllers and without using feedback controllers and results are compared. A sensitivity analysis is further performed to choose the type of controller best suited for the study arterial under given volume conditions.

1.5 Dissertation Organization

This dissertation begun with an overview of the subject. Chapter 2 contains a detailed review of the pertinent literature. A description of the conceptual framework incorporating feedback control in existing adaptive control strategies and elements of the framework is provided in Chapter 3. This chapter also discusses an example of a feedback control system design, the system elements and types of regulators. Chapter 4 illustrates the efficiency of the feedback control systems with the help of a numerical example. It contains the mathematical formulation of traffic flow equations, their application in the arterial under consideration and the performance of the arterial under the effect of two different types of regulators. Chapter 5 presents the simulation of the designed system with MATLAB and the solution process involved. Finally, a summary, conclusions and suggestions for future research are presented in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The use of computers and microprocessors in the field of transportation has been a major breakthrough in achieving an improvement in the performance of traffic control systems. With an increase in traffic volume over the years, conventional strategies used at signalized intersections, like pretimed control, proved to be inefficient in serving high traffic demand. In addition, the complicated demand patterns with respect to time and space also led to higher vehicular delays at pretimed signal systems. Hence, the need of a traffic control system that could adapt to changing traffic conditions was realized.

Several adaptive traffic control strategies such as SCOOT and SCATS have been developed, each aimed at achieving specific objectives with each strategy proven to be efficient under a prespecified set of traffic demand and geometric conditions. As few strategies have been able to work with comparable efficiency for all traffic conditions, a constant effort has been made to develop adaptive control strategies that would work with comparable efficiency under variable traffic conditions.

This chapter includes a brief discussion of various adaptive control techniques and strategies developed to date, followed by an overview of the application of a feedback control strategy to various transportation facilities in order to improve their performance. The chapter further provides a detailed description of three adaptive control strategies: SCOOT, SCATS, and OPAC. The discussion includes the procedures followed by these algorithms to obtain adaptive control in urban arterials.

2.2 History of Adaptive Control

The development of demand responsive or adaptive traffic control systems began in the early 1960s. Earlier systems failed to achieve the objectives of obtaining desired reduction in delays and travel times due to imprecise techniques of forecasting and predicting actual demand from historical data. The first traffic control system, which proved successful over fixed time system in the U.S. was the Urban Traffic Control System (UTCS) developed by the Federal Highway Administration in the 1970s. This is an offline approach that uses manual or computerized techniques to determine signal timing plans to be implemented at different hours of the day. These timing plans are based on minimization of delays and stops.

The first online adaptive system that represented a success was SCOOT *(Hunt, 1982)*, which was closely followed by SCATS *(Lowrie, 1982)*. SCOOT is an adaptive traffic control program that monitors traffic volumes and adapts signal timing in real-time to reflect traffic changes. SCOOT consists of a macroscopic simulation model and a real-time optimization model. The simulation model reads upstream real-time traffic data in the form of Cyclic Flow Profiles (CFPs). These profiles are used to predict queue sizes and detect possible changes in demand to identify an appropriate signal timing plans. Based on these queue estimates, within short fixed intervals of time, the traffic model makes a decision whether to alter the phase, offsets and cycle times of the signal.

Under SCATS, the signal timings are governed by a computer-based control logic with the capability to modify signal timings every cycle as demand fluctuates. The cycle length is selected to maintain a maximum degree of saturation. Phase split selection is based on the degree of saturation. Offset plans are selected from a predetermined library

of offset plans based on the traffic volume in each direction. Thus, the control logic of the system has the capability to modify the signal timings on a cycle-by-cycle basis using traffic flow information collected at the intersection approach stop lines.

Both SCOOT and SCATS are centralized in their operation. The advanced decentralized system OPAC *(Gartner, 1985)* acts on each intersection individually within an arterial. A detailed description of the three systems is provided in sections 2.5, 2.6 and 2.7, respectively. This description forms the basis for incorporating feedback control logic into the existing control systems of SCOOT, SCATS and OPAC. This discussion of offline and online strategies for achieving adaptive traffic control gives an overview of the advancement in developing adaptive traffic control strategies from conventional traffic control strategies.

2.3 Existing Control Strategies

Adaptive control strategies can be divided into four broad categories depending on the techniques used in programming and decision making processes of these strategies. The strategies can be divided into: optimization-based, rule-based, hybrid strategies, and strategies using artificial intelligence. Some of the strategies also use a combination of these strategies. Therefore, these strategies cannot be placed under a specific class, and are generalized as artificial intelligence based strategies in this research.

2.3.1 Optimization-Based Strategies

Optimization-based strategies use computational methods to optimize the total performance of a system in terms of measures of effectiveness such as delays, stops, queue lengths, or a combination of all these factors. Some of the optimization-based

models are OPAC *(Gartner, 1985)*, RHODES (Real-Time, Hierarchical, Optimized, Distributed, and Effective System) *(Head et al., 1992)*, and RTACL (Real-time Traffic Adaptive Control Logic) *(Memon and Bullen, 1996)*. A relatively new development, RT-TRACS (Real Time Traffic Adaptive Control System) has a hierarchical structure to select adaptive systems for fluctuating demand conditions. RT-TRACS serves as a platform for the implementation of a variety of adaptive traffic control algorithms, including new and existing signal timing systems. The system has been designed to select the best adaptive traffic control strategy to respond to current traffic and geometric conditions, with the ability to switch between different strategies to best serve demand fluctuations *(Gartner et al., 2001)*. RT-TRACS works at a distributed or local level.

All of the optimization-based models use techniques such as a dynamic programming approach and genetic algorithms for the optimization. These approaches require local analysis capabilities and communication with signal controllers. In addition, linear programming techniques and queuing theory have also been used to optimize the decision variables like delays or queue lengths to study the dynamics of oversaturated volume conditions.

An optimization-based computer algorithm for intersection and arterial control, OPAC, is considered good for demand responsive, decentralized, flexible coordinated systems. The program makes use of a simplified version of the dynamic programming approach, which is a mathematical technique for the optimization of multistage decision processes. The results obtained were comparable to the results of the dynamic programming approach.

The optimization tool used in some of these strategies is the Genetic Algorithm (GA). Genetic Algorithm is an optimization technique that works in steps based on the principle of genetics which selects the best solution out of a set of available solutions by a stepwise procedure. Abu-Lebdeh and Benekohal *(1998)* used GA to maximize the throughput of traffic from an arterial in oversaturated conditions by dissipating queues on the system links. The algorithm reduced queue lengths and caused better progression of vehicles through a series of intersections. The attractive feature of genetic algorithms is that they efficiently handle wide area problems which involve combinatorial explosion of variables and conditions.

Liu and Kuwahara *(2001)* developed a system wide and real time signal control model, which dealt with oversaturated flow conditions. A dynamic linear programming model was developed using queuing theory analysis and mathematical approximation techniques. The objective of the linear programming problem was to minimize the total delay at an intersection or network under real-time flow conditions subject to constraints of capacity and maximum and minimum departure flow rates. The optimum value of the departure flow rate is suitably converted to the green splits. Finally, under the assumption of point queue, i.e. vehicles are assumed to stack vertically, the value of the signal parameters is fed into the simulation model. The model is made to function as a real time model by inputting the current condition queue length and saturation flow conditions to the formulation model for modifying the splits. Thus, simulation and optimization models were integrated to make the system operate in real time. The model was tested on a hypothetical network of three intersections with preset parameter values. Within each discrete time interval, the model gave the optimal departure flow rates and the split

values of every intersection as the output. This study did not deal with optimization of cycle lengths and offsets.

Another attempt to improve the performance of oversaturated arterials was made by Khatib and Judd *(2001)*. The study involved maximization of throughput from the system thus reducing congestion. The maximization was controlled by constraints at system, zonal and local levels. The objective function was bound by constraints of green time allocation, which was related to the maximum storage capacity of a link. Other constraints were the discharge rate constraints, control of internal queues, which accounts for the storage capacity and initial queue at intersection of interest, and a cycle length constraint. The system used queue management techniques to control queue spillbacks. The linear optimization software LINDO (Linear INteractive, and Discrete Optimizer) was used as an optimization tool in this study. The Performance of the arterial was examined before and after optimization by using CORSIM as the simulation tool. The analysis was performed for an isolated intersection and two arterials of two and five intersections, and results used to understand the effect of the size of an arterial on spillback, total travel time and traffic volume of the system. The simulation results indicated that the operational performance of the arterial was greatly improved by the application of the proposed strategy. An optimization of phase timings and cycle length also maximized the capacity of the arterial.

2.3.2 Rule-Based Strategies

Rule-based strategies compare the performance of a system under study with predetermined thresholds for making short-term control decisions. The decision in these strategies is based on preset rules. The Generically Adaptive Signal Control Algorithm

Prototype (GASCAP) *(Owen, 1998)* is a rule-based approach to adaptive control with three basic elements constituting an algorithm. The elements of GASCAP include: a queue estimation algorithm; a basic set of rules for uncongested conditions using the queue estimates to determine the signal state at an intersection; and a simple algorithm to compute splits and offsets if the intersection is congested. Predefined rules were fixed for demand, progression, urgency, cooperation and safety. Each set of rules submitted a recommendation for the next movement to an event list based on the data obtained from the detector. Each recommendation was analyzed to choose the recommendation with the highest priority. For congested conditions, GASCAP used information from upstream detectors to develop a fixed time signal plan. A comparison of GASCAP against the performance of RHODES, OPAC, and a pretimed baseline strategy for unsaturated, saturated and congested networks during a field test indicated that GASCAP was better adjusted to the stochastic nature of the varying traffic conditions.

2.3.3 Hybrid Strategies

There are certain strategies that are hybrid, i.e. optimization and rule-based, such as TACOS (Traffic Adaptive Control for Over-Saturated intersections Strategy), developed by Li and Prevedouros *(2002)*. Simulation results indicated that TACOS addressed, in part, the problems of existing adaptive control strategies by making use of more reliable information as compared to forecasting models. The information used by TACOS included information for decision making process like optimization of phase sequencing and facilitating detection and response to operational anomalies. The methodology of the research consisted of three blocks. First, the methodology involved the development of an adaptive control strategy to solve problems with existing strategies. Second, the

development of an intersection simulator NIT (NETSIM INTEGRATION, TACOS) to simulate an intersection based on pretimed, actuated, and TACOS strategies took place. Finally, the methodology designed building blocks to evaluate TACOS using NIT.

The objective function of TACOS is based on intersection utilization defined as the throughput of the whole intersection per unit time. Instead of being limited by the cycle lengths and green splits, TACOS chooses any phase among candidate phases and provides a non-fixed green time based on preset rules and optimization objectives. The rules established are used in the optimization procedure. The objective of the optimization is to maximize the throughput from the intersection and the main rule used in determining the phase selection is to make the waiting time of the vehicles at the stop line less than a user-determined time. Simulation testing of TACOS under different volume conditions such as heavy flow, cyclical arrivals, and arrivals with an upstream incident has indicated a significant improvement.

2.3.4 Artificial Intelligence-Based Strategies

Certain strategies, which do not fall into any of the categories described above and imitate human thinking and decision-making processes to achieve adaptive control of traffic, are categorized as strategies using artificial intelligence. These strategies involve the use of a variety of artificial intelligence and mathematical techniques such as fuzzy set logic and neural networks. Artificial intelligence imitates human thinking and decision making and incorporates it into systems with the help of microprocessors and programming techniques. The section below briefly describes these techniques and the level to which these were successful in handing congested conditions.

2.3.4.1 Fuzzy Logic. Fuzzy logic duplicates "human decision making" through the use of fuzzy set theory. The theory involves handling of imprecision in the decision making processes by first fuzzifying the imprecise variables and at the end defuzzifying them under a set of predefined rules to arrive at crisp values as in crisp sets. In fuzzy logic, the degree of membership in a set is represented as a variable value within a range [0, 1] where a "0" is indicative of non-membership while a "1" is indicative of a complete membership. Hamad and Kikuchi *(2002)* proposed a fuzzy inference based procedure to measure congestion in which, the two measures of congestion used were the travel speed rate and the proportion of time traveling at very low speeds (below 5 mph).

2.3.4.2 Neural Networks. Neural networks, which work on artificial intelligence, are suitable for pattern recognition and classification. Artificial neural networks are computing systems, which comprise nodes organized in layers. These nodes are interconnected by links of certain weight. The extent to which they are connected is determined by the weight of the link. The weights are adjusted to minimize the quadratic error between estimated and actual values. Hua and Faghari *(1995)* developed a system, Neural Signal Control System (NSCS), which was designed to identify the best phase sequence out of preexisting patterns. A prediction model called Artificial Neural Network Queue Length (ANNQL) model was integrated with optimization technique to achieve optimal control by Su *(1996)*.

This section on existing adaptive control strategies considered some of the research carried out to handle oversaturated conditions using various techniques and was helpful in identifying control factors and variables which are significant in handling oversaturated traffic conditions. Shortcomings of the existing techniques to manage

queue lengths and conditions of a disturbed state were also researched and a need for better queue management was identified.

2.4 Feedback Control Strategies

2.4.1 Overview

Feedback control has found applications in various fields, including transportation where it is used to control traffic flow within a corridor. These control systems have an advantage of being less sensitive to model inaccuracies and unexpected disturbances. Existing linear time-invariant feedback control systems have mainly been designed with the intent to completely replace the existing traffic control systems for roadway facilities. However, these control systems do not possess elaborate optimization techniques similar to those of existing adaptive control systems. Therefore, it would not be prudent to replace existing adaptive traffic control systems with feedback control traffic systems. However, as proposed in this research, properties of feedback control can be used to improve the performance of existing adaptive traffic control systems and overcome some of their deficiencies.

2.4.2 Existing Models

Feedback control has been used as a tool manage congestion in traffic corridors and freeway sections *(Wang and Papageorgiou, 2000)*. It has also found an application in handling freeway volumes by ramp metering techniques *(Papageorgiou, 1995)*. The following section deals with the application of feedback control strategies in handling over and undersaturated traffic conditions in roadway networks, ramp metering and route

guidance. These applications of feedback control demonstrate the ability of this control to enhance traffic performance conditions.

2.4.2.1 Roadway Networks. Based on dynamic traffic models, a local feedback controller was designed by Wey et al. *(1995)* to achieve intersection control for oversaturated conditions. The study considered traffic flow through an intersection in the general linear state space, with a known initial state and prespecified entry flow. Based on this consideration, a time discrete model was developed to evaluate traffic flow through the intersection with the help of all control variables over a chosen time period subject to specific constraints. Local traffic signal control strategies were used under different traffic conditions to develop the models and design the feedback controllers. For the roadway under study, link i is located upstream of link j. The traffic flow on the j-th road section was modeled by the following set of equations.

$$l_j(k+1) = l_j(k) + d_s(k) - d_{out}^j(k) \tag{2.1}$$

$$d_{out}^j(k) = \min \begin{cases} d_{go}^j(k) \\ d_s^j(k) + l_j(k) \end{cases} \tag{2.2}$$

$$d_{in}^j(k) = \sum_i b_{ij} d_{out}^i(k) \tag{2.3}$$

In the above equations, $d_{in}^j(k)$ is the volume (veh/hr) entering the *j-th* road section during $kT_s \le t \le (k+1)T_s$, $d_s^j(k)$ is the volume which arrives at the end of the waiting queue or at the stop-line, $d_{out}^j(k)$ is the leaving volume at the downstream end of the *j-th* road section, $d_{go}^j(k)$ is the capacity of the green portion of the light, $l_j(k)T_s$ is the number of vehicles in the queue and b_{ij} denotes the percentage of volumes leaving road section i and entering road section j.

The state space equations used to describe the system are as follows:

$$l_1(k+1) = l_1(k) + [d_1(k) - r_1(k)]t_c \tag{2.4}$$

$$l_2(k+1) = l_2(k) + \left[d_2(k) - S_2\left(1 - \frac{L}{t_c}\right) + \frac{S_2}{S_1} \times r_1(k)\right]t_c \tag{2.5}$$

$$r_i = S_i \times \frac{g_i}{t_c} \tag{2.6}$$

The state variable is the queue length, i.e. $l_1(k)$, the controllable input variable is $r_1(k)$, the disturbance is $d_1(k)$, and k is the time index. Queue lengths for competing flows are denoted by $l_1(k)$ and $l_2(k)$, $d_1(k)$ and $d_2(k)$ denote the arrival rates of vehicles, S_1 and S_2 denote the saturation flow rates for competing flows, r_i represents the capacity flow for the green portion of the light, L stands for lost time and t_c is the cycle length. This algorithm calculates queue lengths for the main and cross streets separately.

$$r_1(k)t_c = r_1(k-1)t_c - E.(err(k)) \tag{2.7}$$

where

$$err(k) = l^d - l(k) \tag{2.8}$$

The objective is to minimize the error term as represented in equation (2.8). The error is determined as the difference between the desired and existing queue lengths. These lengths are calculated separately for the main and cross streets. The formulation proposed in this study was not applicable to real time control systems since a real time application requires the development of an adaptive prediction algorithm and optimal

control policy. Also, this research did not consider any algorithm that could predict the future fluctuations in traffic volumes and accordingly adapt to the varying volume conditions.

In a subsequent research project, Wey et al. *(1999)* described the traffic flow and dynamic reduction of queues using a linear model with linear capacity constraints for both road links and intersections. This study extended the previous work by introducing an optimization routine to the existing state-space model. The optimization routine minimized the total delay within the network and was applied to real time operations. The optimization was carried out through mixed integer linear programming (MILP) to achieve minimization of delays and to permit queue build up to a predetermined upper bound. Finally, an efficient special-purpose version of network simplex was developed to solve the programming problem that provided solutions faster than the other optimization programs, such as LINDO, which use standard simplex and branch and bound techniques. The model had the capability to be incorporated in a sensor-based as well as a feedback control framework.

This approach was tested by Wey and Jayakrishnan *(2001)* by comparing it with conventional signal timing models, using the microscopic simulation model TRAF-NETSIM. To make the system function in real time, NETSIM generated the simulated output data of arrival flows every 5 seconds, which was fed into the optimization control model. The optimization model in turn generated signal control data that was once again fed into NETSIM to obtain a summary of traffic measures of effectiveness. Vehicle delay and link-specific delays were chosen as a MOE to compare the relative efficiency of actuated control, optimization control, and rolling horizon control. It was found that the

optimization traffic signal control algorithm showed a significant improvement over the actuated control timing plans and the existing control timing plan of 2.29% and 11.04%, respectively. The proposed algorithm was found to perform well under moderate or heavy traffic volumes.

Another attempt to use feedback control to manage queue lengths on urban roadways was made by Shimizu and Kita *(1996)*. First a congestion mechanism was developed to identify the level of congestion based on general traffic flow equations. The traffic flow equations were developed to maintain the same level of congestion on the main and cross streets. The signal control system was determined using a nonlinear time-varying discrete dynamic system. An algorithm was developed to control three parameters consisting of cycle length, green split and offset in order to minimize the sum of queue lengths on the arterial. A time step was considered equal to the cycle length. The volume balance at an intersection is determined using the following equation:

$$x_e(i,j,m,k) = x_e(i,j,m,k-1) + x_i(i,j,m,k) - x_o(i,j,m,k) \qquad (2.9)$$

where i and j denote the location of each signalized intersection with i=1,2... representing main streets and j=1,2... representing cross streets. Also, m =1,2,3,4 represents the moving direction of vehicles respectively with m=1,2 representing east bound and west bound movements while m=3,4 representing north and south bound movements respectively. The incoming volume is represented by $x_i(i,j,m,k)$ while the outgoing volume $x_o(i,j,m,k)$ is controlled by control parameters cycle length, green split and offset. The state variable is represented as $x_e(i,j,m,k)$ which is the queue length to be served in terms of vehicles at time index k. The state variable is further converted to a

queue length by multiplying it with a transformation factor $l_m(i,j,m,k)$. The signal control system was represented by the following set of state-space equations:

$$x_e(i,j,m,k) = x_e(i,j,m,k-1) + x_i(i,j,m,k) - u(i,j,m,k)$$ (2.10)

$$y_c(i,j,m,k) = l_m(i,j,m,k)x_e(i,j,m,k)$$ (2.11)

where y_c denotes queue length and $u(i,j,m,k)$ represents the control input.

The excess volumes were calculated by the following expression

$$x_e(i,j,m,k) = x_i(i,j,m,k) - x_0(i,j,m,k)$$ (2.12)

The output from the system was calculated as

$$x_o(i,j,m,k) = \xi(i,j,m,k)c_x(i,j,m,k)$$ (2.13)

where $\xi(i,j,m,k)$ attains the value of effective green time. The equation for the congestion length, i.e. the output, was expressed as a function of the input volume of the vehicles expressed in terms of queue lengths using $l_m(i, j, m, k)$ as the transformation factor. The transformation factor is used to convert the number of vehicles to a corresponding queue length.

The objective of the signal control system was to find the control input value which would minimize the performance criterion $J_n(k)$. The performance criterion, $J_n(k)$, represents the summation of errors along the arterial. This can be represented as

$$J_n(k) = \sum_{i=1}^{L}\sum_{j=1}^{N}\sum_{m=1}^{4} g(i,j,m,k)$$ (2.14)

where L and N denote number of the signalized intersections. The function g(i,j,m,k) and control error e(i,j,m,k) are defined by

$$g(i,j,m,k) \equiv \begin{cases} 0, when & e(i,j,m,k) \geq 0 \\ |e(i,j,m,k)|, when & e(i,j,m,k) < 0 \end{cases} \qquad (2.15)$$

$$e(i,j,m,k) \overset{\Delta}{=} l_r(i,j,m,k) - y_c(i,j,m,k) \qquad (2.16)$$

where the control error represents the difference in the excess volumes or queue length for two subsequent time periods. Green times, green splits and cycle lengths were evaluated and adjusted for all directions based on the relationship among signal control parameters at each intersection of the arterial.

As an extension to the study described above, Shimizu et al. *(1999)* analyzed the control of queue lengths for a one-way traffic network from a deterministic control viewpoint. Two signal control algorithms were presented in this study to control queue lengths on arterials and roadway networks formed by a combination of several arterials. The algorithms developed for the arterial were the "balanced control algorithm" and the "priority control algorithm" while the algorithm developed for roadway networks was the "network control algorithm". As described above, the performance criterion to be minimized was the errors in the system. Three signal control parameters, i.e. cycle length, greens-splits and offsets, were stepwise evaluated to minimize the sum of the control error function for the traffic network. For the network control algorithm, the three control parameters were evaluated for the complete network in order to reduce the overall control error for the network.

The feedback control system was used to evaluate the maximum value of the cycle length for a given demand and then, the other two control parameters were evaluated to minimize the performance criteria using the balance control algorithm. The evaluation of offsets was performed to optimize the performance criterion which is the summation of errors for all approaches of an intersection and for all intersections on the study arterial. The balanced control algorithm was tested at four consecutive intersections of an arterial while the network control algorithm was simulated at twelve signalized intersections. The simulation results indicated that the proposed signal control algorithms, with an implementation of feedback control, satisfied all performance criteria on the arterial and traffic network. Simulation results confirmed that cycle lengths were controlled in a wide range according to the large variation in demand. Green splits and offsets were satisfactorily controlled to minimize the performance criteria. A comparison of the measured and calculated values of congestion on the arterial indicated that little or no congestion existed on some of the links after the implementation of the proposed algorithm.

The same study was extended by Shimizu and Mashiba *(2000)* to signal control for two-way arterials. Volume balance was maintained for each lane at each signalized intersection of the traffic network and was represented using difference equations. The difference equations used excess volumes as a function of incoming and outgoing volumes. Within an arterial, feedback control of queue lengths was applied. In the balance control algorithm, the green time was controlled to minimize the performance criterion which is the summation of error on all approaches to an intersection for all intersections on the study arterial. The control algorithm for congestion lengths was

simulated at four signalized intersections in Fukuyama City, Japan. The cycle lengths were controlled for the simulated intersections according to the variation of incoming volumes. A comparison of the cycle lengths between the simulation value and the measured values indicated that cycle lengths were adaptively controlled by the balance control algorithm. It was found that the queue lengths were controlled to become nearly zero at the four signalized intersections.

Papageorgiou *(1995)* presented an optimal integrated control strategy for a number of control elements including signal control, ramp metering, motorway-to-motorway control, VMS control, and route guidance to achieve queue management. The proposed algorithm which is applicable offline as well as in real time, is based on current traffic measurements. Linear optimal control problems were modeled for each traffic corridor using a store-and-forward approach. The store-and-forward approach is a mathematical modeling approach particularly suitable for saturated conditions and provides gating to protect downstream areas from oversaturation. An objective function was formulated with constraints on the effective green times and maximum allowable queue lengths in order to minimize the total delays. The objective function was further extended for networks with an objective to minimize the total travel time within the network. Thus, the control problem was treated as a linear optimal-control problem. An algorithm *(Banos and Papageorgiou, 1995)* was used in which each optimization run was based on updating the initial condition that includes the current values of all state variables of all network queues.

It was attempted to make the system compatible for both saturated and undersaturated traffic conditions, but it was found that the store-and-forward approach

led to a medium–term control plan in highway networks mainly under saturated traffic conditions. The approach proved to be ineffective for undersaturated conditions. According to a store and forward approach, vehicles experience constant travel times along a link and are stored at the end of the link if the corresponding inflow is higher than the outflow. The outflow from a link is forwarded according to the nature of the corresponding node and the control measures applied.

The store and forward approach for queue management was used for an urban road link between two adjacent intersections as shown in Figure 2.1. As shown in the figure, $x_i(k)$ represents queue length at time index k, $x_i(k+1)$ is queue length at time index $k+1$, $\kappa_i T$ is the constant travel time for non critical travel densities, $q_i(k)$ and $u_i(k)$ are the inflow and outflow respectively of link i over a period [kT, (k+1)T], $d_i(k)$ is the demand flow, $x_i(k)$ represents the queue length, t_{ij} are turning movement rates, $s_i(k)$ is the exit flow, and t_{io} exit rates for turning movements. The demand is assumed known, and the exit flows can be represented as follows:

$$s_i(k) = t_{io}q_i(k) \qquad (2.16)$$

The state space equation representing queue formation within a link is stated as follows:

$$x_i(k+1) = x_i(k) + T[(1 - t_{io})q_i(k - \kappa_i) + d_i(k) - u_i(k)] \qquad (2.17)$$

M

N

u_i

q_i

Link l_i

Link l_i

s_i   d_i
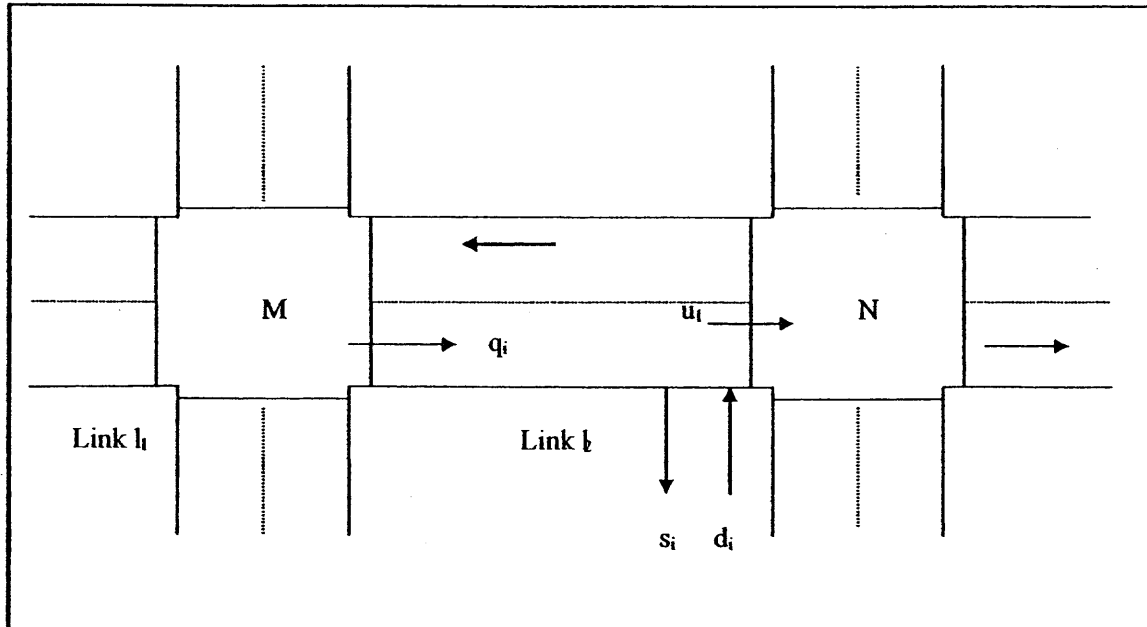
Figure 2.1 Traffic flow at intersections.

The equation is an expression of the queue length over a time period [kT, (k+1)T] where

T is the sample time interval and k=1,2... is a discrete time index. Thus the state variable

x_i is expressed as the summation of existing queue length from the previous time period

[(k-1)T, kT], the difference between the inflow and the outflow of the main link i over

time period [kT, (k+1)T], and the difference between inflow and outflow of the cross

street. The term $(k - \kappa_i)$ represents the effective time during which the queue build up

takes place.

The above state space equation describes the dynamics of link i expressed as a

combination of a constant travel time equal to $\kappa_i T$, and a delay associated with queuing.

It was assumed that all the movements (through, right and left turning) of the incoming

link receive the right of way simultaneously. Therefore, the state space equation takes

into account inflow into the links, outflow from the link, inflow from the cross street and

the turning movement into the cross street. The inflow to link i can be expressed as a

summation of through volumes and turning volumes into the intersection j. This is expressed as follows:

$$q_i(k) = \sum_{j \in I_{in}} t_{ji} u_j(k) \qquad i \in I_N, j \in O_N \qquad (2.18)$$

The objective of the procedure is to minimize the total delay at both the main and the cross street of an intersection. The algorithm considers the turning movements from and into streets while calculating the volumes on a link. For time period T, chosen not less than the cycle time c, an average value of the link outflow $u_i(k)$ is obtained, which is given by

$$u_i(k) = S_i G_i(k) / c \qquad (2.19)$$

where, S_i represents saturation flow, c represents cycle length and G_i represents effective green time for each link i. The effective green time G_i is introduced separately for different intersections in order for the system to be compatible for both over and under saturated conditions. The green time can be varied depending on the level of saturation at different intersections. Here, time period T is taken equal to the fixed cycle length for different intersections. Other cases are analyzed which considered situations in which the time period was less than or more than the cycle length. Real time implementation of this method requires availability of historical or real time forecasts of traffic demand and availability of queue measurements.

2.4.2.2 Ramp Metering. A traffic responsive strategy, ALINEA, for ramp metering based on a feedback structure was derived by Papageorgiou et al *(1991)* using automatic control methods. The function of the control strategy, ALINEA, is to control the input volumes from the ramp in order to maintain a desired occupancy downstream of the on-

ramps in case of uncontrollable through traffic on the freeway. The first step involves the derivation of the feedback law, followed by the design of the regulator. Site conditions and the traffic flow process are as shown in Figure 2.2, where site 1 is assumed to be situated just upstream of the on-ramp, site 2 is situated downstream of the on-ramp, at a distance δ from site 1. The function of ALINEA is to control the input volumes from the ramp in order to maintain a desired occupancy at site 2 when through traffic from site 1 is an uncontrolled variable. Thus, ALINEA determines the metering rate of the ramp in order to achieve the desired performance at the downstream section of the ramp.
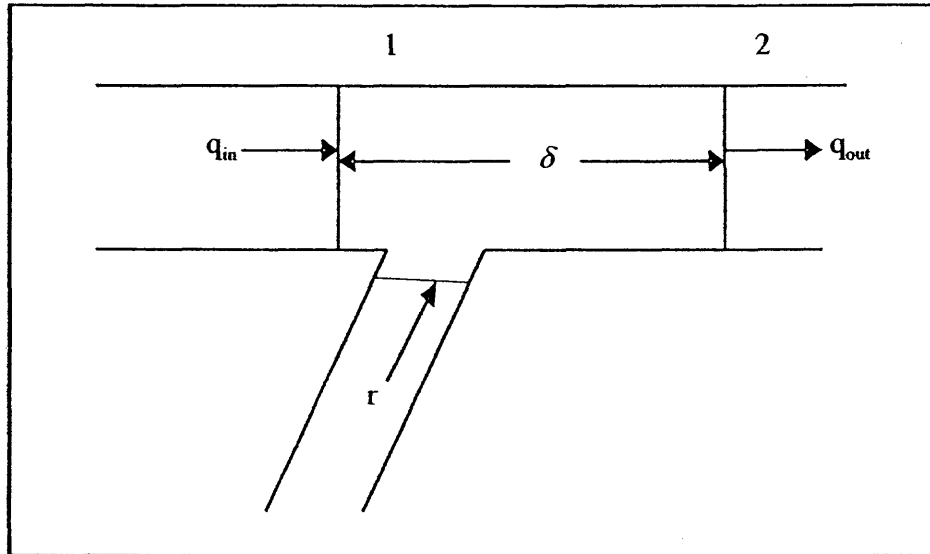
Figure 2.2 Traffic flow process in ALINEA.

Depending upon whether the desired occupancy is lower or higher than the actual occupancy for the current time period, the feedback control system ALINEA decides to increase or decrease the ramp-metering rate with respect to the metering rate provided in the previous time period. The occupancy at a time period k is given by the equation

$$\Delta o_{out}(k+1) = \beta \Delta o_{out}(k) + \left[(1-\beta)/\hat{Q}' \right] \times \left[\Delta q_{in}(k) - \Delta r(k) \right] \qquad (2.20)$$

$$\hat{Q}' = dQ(\hat{o})/do_{out} \qquad (2.21)$$

where k = 0,1,2,… represent the sample time index. Thus $o_{out}(k)$ is the occupancy at time kT. The time discretized values of disturbance q_{in} and control input r are denoted by $\Delta q_{in}(k)$ and $\Delta r(k)$ respectively and are assumed to be constant during the time interval [(k-1)T, kT]. The constant parameter β results from a discretization procedure and can be neglected in certain cases when the ratio δ/T is sufficiently small and δ represents traffic density. Thus taking $\beta = 0$ for very small time intervals T, leads to

$$\Delta o_{out}(k+1) = [\Delta q_{in}(k) + \Delta r(k)]/\hat{Q}' \qquad (2.22)$$

A feedback law for the above expression can be given by the regulator

$$r(k) = r(k-1) - K_R\left[\hat{o} - o_{out}(k)\right] \qquad (2.23)$$

where, K_R is a constant regulator parameter. Application of this regulator to the original process model of equation (2.20) gives a closed loop system that has low transient behavior. An improvement to the system can be achieved by applying a proportional-plus-integral feedback law given as:

$$r(k) = r(k-1) - K_R\left[\hat{o} - o_{out}(k)\right] - K_P[o_{out}(k) - o_{out}(k-1)] \qquad (2.24)$$

where, K_P is a constant and positive regulator parameter value. For a constant upstream traffic volume q_{in}, the above feedback law leads to $o_{out} = \hat{o}$ in the steady state. Thus, the feedback law leads occupancy to its desired level irrespective of the constant upstream traffic volume.

A field test and implementation of ALINEA was made and it is currently operational at on-ramps in Paris. This ramp metering strategy is known to have advantages over other known strategies with respect to its simpler algorithm, minimal requirement of real time measurements and adjustability to particular traffic conditions. It also has an ability to be embedded in a coordinated on-ramp control system and has the flexibility of modification in case of changing operational requirements. Due to feedback properties, ALINEA is more robust with respect to inaccuracies and disturbances in input variables and also has a strong theoretical support by automatic control theory.

A comparison was made between the performance of ALINEA and five other feed-forward control strategies on a ramp in southern Paris by Hadj-Salem et al. *(1990)*. The first feed-forward control technique was static control which involves restricting geometrically the end of the ramp so as to force vehicles entering the mainstream in one lane. The second strategy was a fixed-time control, while the third strategy, called Demand-Capacity strategy, was based on comparing the demand on the upstream of the merge area with the capacity of the bottleneck downstream of the merge area. The fourth strategy tested was the percent-occupancy strategy, which was essentially based on the philosophy of the Demand-Capacity strategy, but the former used occupancy measurements for estimating upstream demand and this strategy uses capacity measurements for ramp metering. Demand capacity INRETS was the fifth strategy tested which utilized measurements from three down stream detectors in order to have a better estimation of the degree of congestion. These five feed forward techniques were tested and compared to ALINEA, which made use of feedback theory. The six strategies were compared in terms of total travel time, number of vehicles served, congestion duration

and estimated number of diverted vehicles for free flow demand conditions. Field results showed that ALINEA led to a 3% increase in the number of vehicles served, 55% reduction in the time spent in congestion, 19% reduction in the mean travel time and it was also found to be superior in terms of diverting less traffic.

Oh and Sisiopiku *(1998)* developed a modified version of ALINEA with upstream monitoring as the basis of ramp metering. The modified model was formulated to overcome some of the shortcoming of the ALINEA model. The fault with ALINEA was that after implementation of the downstream optimization and several periods of ramp metering, the downstream occupancy would be set to an optimum value, whereas the volume conditions on the upstream of ramp may oscillate between congested and uncongested conditions. When the downstream functions at an optimum occupancy, congestion could still exist on upstream of the on-ramp. Studies have indicated the significance of field location in the resulting speed-flow-density relationships. These studies also show that upstream of the optimized bottlenecks is considered to be the worst sections in terms of speed and occupancies under congested conditions.

Secondly, it was difficult to determine the best detector locations for downstream control models. The modification of ALINEA included firstly, parametric studies in terms of traffic characteristics and conditions. These studies involved obtaining data at the upstream and downstream of the on-ramp and obtaining the volume-occupancy relation curves. The data were used to obtain the occupancy relation parameter between occupancies upstream and downstream of the on-ramp. Secondly, model generalization, construction and validation were performed by using field data to measure the efficiency of the new model in terms of standard error.

Finally, a feedback operation study was performed using simulation to demonstrate the operational effectiveness of modified ALINEA and to reflect the interactive operation between ramp volumes and occupancies at the ramp merge under the effect of feedback control. In the model modification, a relationship was established between upstream and downstream occupancy of the on-ramp which also provided a relationship between occupancies in bottleneck cases. An added feature in the modified ALINEA was that it accounted for the time lag between ramp metering discharge and downstream response. The optimum time lag was obtained using linear regression technique for different volume conditions.

These added features incorporated into ALINEA model provided a feedback control system, which was tested using Monte Carlo (MC) simulation. The parameter such as those relating the occupancies at upstream and downstream, and regulator parameters of the controller were taken as constants for the purpose of simulation since they are system specific. Data for the study were obtained from Michigan Intelligent Transportation System (MITS), which operates freeways within Detroit metropolitan area. The measures of effectiveness used to compare ALINEA and MALINEA were occupancy, recovery speed, coefficient of variation, ramp discharge rate, total non-delay travel time, accumulated mainline delay, accumulated ramp delay, total travel time and total distance traveled. The results of field-testing of the model show that the new strategies showed improvement in MOEs. However, the results showed an increase in accumulated ramp delay, with a reduction in all other delays. Thus it is required that more reliable field results are obtained, and also, there is a necessity of testing a speed-based model.

The discussion on ALINEA and MALINEA helped to understand the effectiveness of different types of controllers in achieving desired performance from a feedback control system. It also proved the benefits of feedback control over feed-forward control. The study of ALINEA and MALINEA described the concept of deciding appropriate performance criteria and measures of effectiveness, which in this case was the optimized occupancy at the downstream of the on-ramp.

2.4.2.3 Route Guidance.   Palvis and Papageorgiou *(1999)* developed a feedback route guidance strategy for roadway networks. The study implemented a feedback control strategy to multi-origin multi-destination corridor networks. Two traffic networks, small and large, of similar mesh structure were analyzed. The objective of the strategy was to split the total traffic flow into three routes using a controller in order to achieve equal travel times on all routes. Regulators such as bang-bang, P and PI regulators were made to function under a variable set of demands and incident conditions. Simulation was performed for all the above mentioned conditions in order to test the behavior of the designed decentralized feedback route guidance regulator. It was found that the decentralized control laws successfully improved the system performance in terms of total time spent in the system.

The feedback route guidance strategy that was developed for meshed traffic networks to establish equal travel times on alternative routes leading to a particular destination, despite the impact of disturbances in the form of incidents, weather conditions, varying demand, road works, compliance etc. is described below.

In the absence of route guidance, the P-regulator is given as:

$$\beta_{nj}^{m}(k) = \beta_{nj}^{mN} - K_{P}\Delta t_{m\mu}^{j}(k), \qquad \beta_{nj}^{m}(k) \in [0,1] \tag{2.25}$$

where $\beta_{nj}^m(k)$ denotes constant splitting rates of traffic for destination j via route m at the bifurcation node n, K_P denotes the P - regulator's parameter, and $\Delta t_{m\mu}^j$ denotes the difference in the travel times to destination j via route m and μ. Using control theory, an increase in the parameter K_P leads to the oscillatory behavior of the system which represents instability of system. If instead of a P-regulator, a PI regulator is provided, then

$$\beta_{nj}^m(k) = \beta_{njm}^m(k-1) - K_P\left[\Delta t_{m,\mu}^j(k) - \Delta t_{m\mu}^j(k-1)\right] - K_I \Delta t_{m\mu}^j(k) \qquad (2.26)$$

$$\beta_{nj}^m(k) \in [0,1]$$

where K_P and K_I are regulator parameters. The values of the regulator parameters are found by a trial and error method which adjusts the values of regulator parameters in such a way that the actual output value traces the desired set value of the output.

In the presence of route guidance and driver compliance to the guidance, the following shows the splitting rate:

$$B_{nj}^m(k) = \varepsilon\beta_{nj}^m + (1-\varepsilon)\beta_{nj}^{mN} \qquad (2.27)$$

where $0 \le \varepsilon \le 1$ is the compliance rate. Incorporation of the factor ε in the above model does not change the general structure of the P or PI regulator. The compliance rate only modifies the values of regulator parameters and thus can be considered as an external disturbance to the control loop. The implementation of the system within two network systems proved that in case of an extensive network, the overall strategy is not required to be redesigned. Both the P and PI regulators were found to be versatile in terms of handling different situations of O-D, incidents, and driver compliance without involving the complicacy of real time predictions.

Kotsialos et al. *(2002)* further extended the previous work by considering separate macroscopic traffic flow models for non-destination and destination oriented trips. In non-destination oriented trips, the maximum outflow was determined by the demand raised on the two ramps under consideration. For destination-oriented trips where a traffic assignment is required, splitting rates need to be considered at bifurcations. The objective was to minimize the total time spent in the system, which is equal to the summation of total travel time and waiting time. The objective function was expressed as a discrete-time optimal control problem in state space form and a feasible-direction algorithm for numerical solution was provided. The proposed model was tested on a two ramp hypothetical test network for different scenarios, which were: no control, coordinated ramp metering, route guidance, integrated control, integrated control with congested initial state and integrated control with maximum queue constraint. It was found that the strategy was fair with respect to both on-ramps without causing any significant handicap to the traffic volume served by any of the two controlled ramps nor does it favor any one of them at the expense of the other.

This section discussed the details of previous methodologies using control system. The review helped to develop an understanding of the concept of state-space, disturbances in the system and the effect of the types of controllers on the system to achieve the desired state. It also provided insight regarding the choice of control variables and appropriate performance measures for evaluation of systems. It was found that all the strategies discussed above using the concept of feedback control were aimed to replace the existing control strategies for the facilities under consideration. The proposed strategy aims at overcoming some of the shortcomings of the existing adaptive control techniques

with the help of feedback control, while still using the diversity of the existing adaptive control systems.

2.5 Split Cycle Offset Optimization Technique (SCOOT)

SCOOT has been identified as a leading adaptive control strategy which has found an application in several major cities. The SCOOT system was originally developed in the United Kingdom at Transport and Road Research Laboratory (TRRL) in the 1970s. The system continued to evolve over time to accommodate complex traffic conditions *(Hunt, 1986)*. Additional functions have been added to SCOOT to accommodate congestion management during incidents with signal timing plans overriding the SCOOT optimizer plans for emergency or transit vehicles. A gating facility to reduce congestion at bottlenecks, improved data structures for left-turn bays, and bicycle logic, to name a few, have also been added.

2.5.1 Principles of SCOOT

The primary objective of the basic model of SCOOT is to minimize the sum of the average queues in a network. The SCOOT urban traffic control system achieves this objective by following a series of functions. This section provides a description of the procedures followed by SCOOT starting from data collection, data processing, traffic flow model and optimization procedures. The performance index (PI) in SCOOT is expressed in terms of the sum of the average queues in a network. The physical significance of the PI is in terms of delays which can be translated to related cost. The PI in SCOOT also takes into account the number of times vehicles have to stop. In general,

signal timing plans that minimize queues also reduce stops. However, when the objective is to minimize stops, the model has a tendency to favor longer cycle times.

The three key functions used by SCOOT to obtain adaptive control at arterials are: measuring cyclic flow profiles (CFPs); running the online traffic flow model based on these CFPs; and incrementally optimizing the cycle time, splits and offsets. The first step represents the measurements of data and calculation of CFPs. Cyclic flow profiles are an online measurement of vehicle flow by vehicle sensors installed upstream from each signal stop line. A CFP is defined as a measure of the average one-way flow of vehicles past any chosen point on the road during each part of the cycle time of the upstream signal. In order to measure CFPs, the sensors are located well upstream of stop lines, preferably just downstream from the previous junction. This location facilitates the earliest possible prediction of arrivals at the downstream stop line. It also allows an early warning of gridlock, which can occur if capacity is lost at the previous junction. Thus, CFPs record platoons of vehicles at successive steps within the cycle. The CFPs contain information on the demand for green time as well as the information needed for the proper coordination of signal timing plans of adjacent intersections in an arterial.

In the second step, based on the CFP, a model is used to estimate how many vehicles will reach downstream signals during the red time. The information for this calculation is provided in the online traffic model. Both TRANSYT and SCOOT contain similar traffic models that are able to estimate queue length. These traffic models are used by optimizers to evaluate alternative signal timing plans. In SCOOT, it is assumed that traffic platoons travel at a known cruising speed with some dispersion and the queues are discharged at a saturation flow rate during green time. The saturation flow rate is

known and constant for each signal stop-line. Based on the state of the traffic signal and a preset saturation flow rate, the length and the back of the queue are estimated. Hence, both the size of the queue and how long it takes to clear it can be calculated and the effects of alterations in the signal timings predicted. Figure 2.3 provides a detailed explanation of the measurement of CFPs and obtaining queue lengths from these flow profiles. The estimations of growth and clearance of queues are updated every 4 seconds.

The third step SCOOT carries out is the optimization of splits, offsets, and cycle times to optimize the performance criteria. The SCOOT online system has a set of fixed signal timing plans which are altered by the optimizers to meet the requirements of changing demand conditions. Thus, during the third step, calculation of the coordination plan to respond to new traffic situations through a series of small frequent increments is calculated.

Operation of SCOOT takes place on a regional basis. Each region has its common cycle time to maintain coordination. For each region, the model calculates the degree of saturation for all its nodes. Next it identifies the most critical node and calculates the optimum cycle length. The critical node may, and frequently does, change over time. All nodes are assigned the optimum cycle time and changes are imposed to offsets and splits with the help of optimizers. The model also provides an option of double cycling where the model has a flexibility to assign half the optimum cycle time to any node if found to be advantageous.
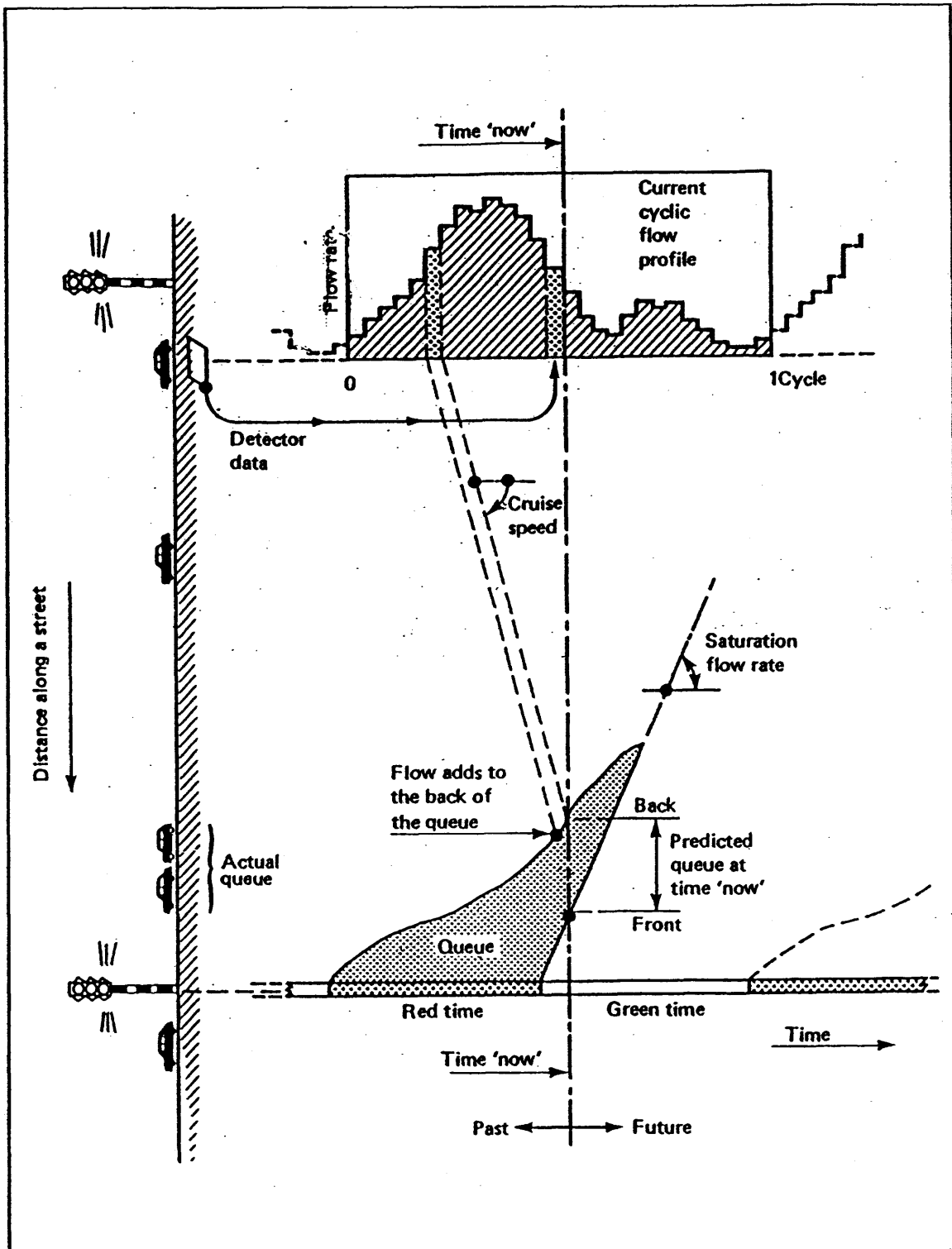
Figure 2.3 Concept of Cyclic Flow Profiles and queue management in SCOOT.

2.5.2   SCOOT Optimizers

Three SCOOT optimizers are the cycle time optimizer, split optimizer and the offset optimizer. The cycle time optimizer can vary the cycle of each sub-area in increments of a few seconds at intervals of not less than 1½ minutes. Each sub-area's cycle length is varied independently from other sub-areas between preset upper and lower bounds. The cycle time is varied by SCOOT to ensure that the most heavily loaded junction operates, if possible, at a maximum degree of saturation of 90 percent. The SCOOT optimizer accounts for conditions when vehicles queue over the sensors.

The offset optimizer operates on each junction's cycle lengths. The information in the CFP is used to estimate whether or not an alteration to the offset will improve the overall traffic progression on those streets which are immediately upstream or downstream of the junction. Therefore, once every cycle, the offset optimizer assesses whether altering the current offset by 4 seconds earlier or later can reduce the PI on streets around each junction. The split optimizer implements the alteration that will minimize the degree of saturation on the approaches to that junction. Any alteration made to green duration is temporary and for each temporary alteration, a smaller permanent alteration is made to the signal plan. Favorable split and offsets that minimize the performance index are implemented immediately. The performance index uses delay, stops and level of congestion as parameters of system performance.
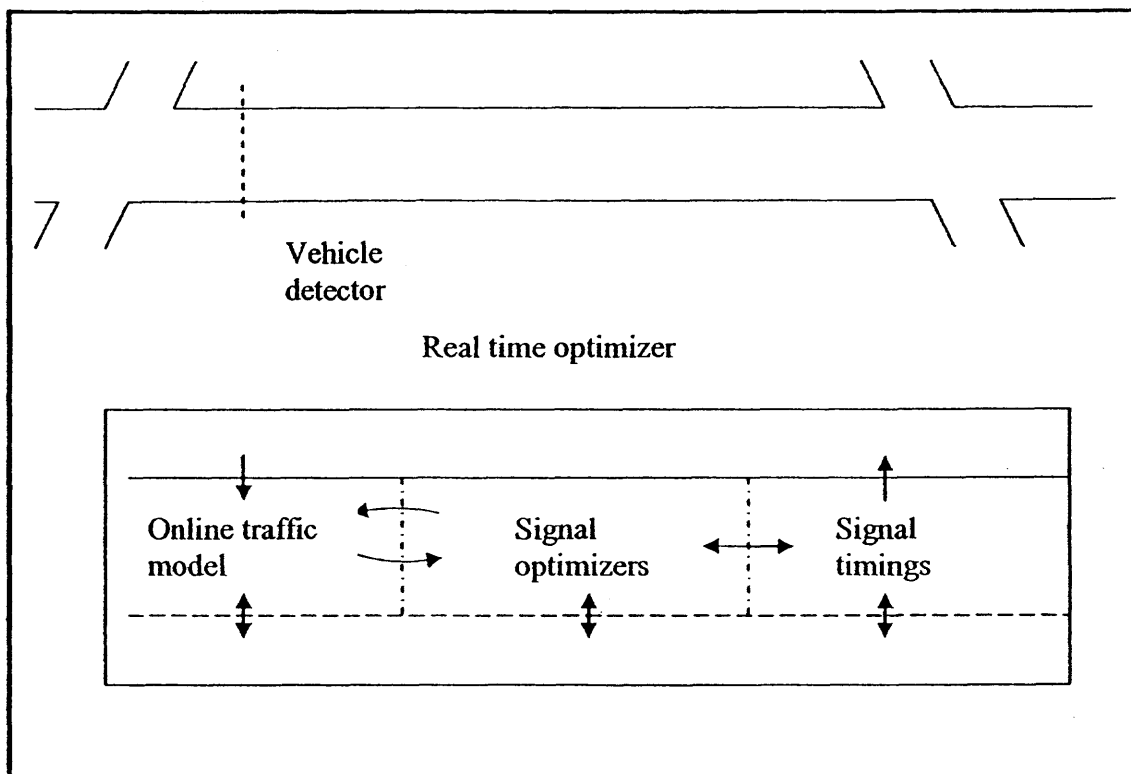
Figure 2.4 Flow of information in a SCOOT urban traffic control system.

To better adapt to the latest traffic flow conditions recorded by the CFPs, the SCOOT optimizer uses an "elastic" coordination plan. Thus the plan represents optimized splits, offsets, and cycle times. According to the elastic coordination plan, a few seconds before every phase change, the SCOOT split optimizer calculates whether it is better to advance or retard the scheduled change by up to 4 seconds, or to leave it unchanged. Then once every cycle, the offset optimizer assesses whether changing the offset of that junction by up to 4 seconds either way can reduce the PI on the streets around each junction.

Upon finding the optimal signal plan for the arterial, SCOOT implements the signal plan with proper offsets in order to obtain coordination amongst adjacent pairs of intersections. Thus, the entire operation of SCOOT relies on the accuracy of data

collection and evaluation of the traffic and queue condition by the traffic model. The model does not undergo a verification or evaluation process to find out the effectiveness of the implemented plan. The traffic model calculates the PI associated with the signal plan before the implementation phase. The flow of information within SCOOT as described above has been presented in Figure 2.4.

Provisions for accommodating special traffic conditions like incident or emergency conditions when the optimized functions are changed to manage the special conditions are also provided within SCOOT. In special cases of oversaturation, an overriding plan changes the objective to maximize throughput instead of minimizing queues.

Experiments with SCOOT have led to the belief amongst researchers that SCOOT's delay minimization objectives may not be the most effective technique for controlling networks when there is a high level of congestion. Hence, research was carried out (Wood et al., 1994) to obtain an understanding of the mechanism by which congestion spreads within networks and to investigate situations and circumstances where SCOOT in its normal mode of operation looses its effectiveness. A wide range of networks were studied under a range of flow conditions to determine the performance of SCOOT under congested conditions. The study also determined traffic conditions under which the effectiveness of SCOOT is reduced or lost. The conditions under which SCOOT provided suboptimal results were when the main road link length is short and the saturation flow of the main road is greater than that of the side road. In this situation, spillback on one link can lead to the blocking of a number of other links causing congestion to spread. To identify the need to introduce a special congestion strategy

referred to as a congestion indicator, Wasted Capacity, was derived. Wasted capacity is given as the minimum out of the values of lost capacity and the queue at the end of the green time. Wasted capacity at a node is the sum of the wasted capacities for the upstream link of that node. To implement the concept, first the wasted capacity is measured on the links in an area. This is followed by the formation of trees of nodes where wasted capacity occurs, and for each tree, critical links and nodes are determined. Depending on whether the critical link is short or the offset on the link is unstable, appropriate measures are taken by the system.

2.6 Sydney Coordinated Adaptive Traffic System (SCATS)

The Department of Main Roads, N.S.W. developed SCATS in the early 1970s to obtain coordination of a large number of Sydney's traffic signals situated on the main arterial roads *(Lowrie, 1982)*. The hierarchy of a SCATS system includes regional computers at the highest level, which maintain a traffic responsive control of local controllers. The local controllers are grouped into 'systems' and 'subsystems'. Subsystems are the basic unit of strategic control within SCATS and may typically consist of one to ten intersections. Subsystems may or may not interact with each other depending upon the traffic requirements. Systems do not interact amongst each other since they are geographically separated. A central computer is also provided for centralized monitoring of system performance and equipment status.

2.6.1 Signal Plan for SCATS

Four green split plans are provide for each intersection. The strategic control algorithm makes a selection of the split plan at the subsystem level depending upon the demand at a critical intersection within the subsystem. Five internal background offset plans are provided to determine the offsets between intersections within the sub-system, and five external offset plans are provided for linking adjacent subsystems. The basic traffic data measured by SCATS for strategic control is a measure analogous to the degree of saturation on each approach. The local controller collects flow and occupancy data from the detectors during the green phase of the approach. After pre-processing and filtering, the data is sent to a regional computer to calculate the SCATS "degree of saturation" (DS). Degree of saturation in SCATS is defined as the ratio of the effectively used green time to the total available green time on the approach. The effectively used green time is the length of green that would be just sufficient to discharge the same platoon of vehicles if they would have been traveling at optimum headways in saturation flow conditions.

2.6.1.1 Cycle Length. The subsystem cycle length is a function of the highest DS measured in the sub-system during the previous cycle. The change in the cycle length, which is limited to ±6 s, is given by

$$C' = 60[DS\text{-}f(C)] \qquad (2.28)$$

where C' is the change in cycle length for the following cycle and C is the cycle length of the current cycle, which is a function of the highest DS for the current cycle. Thus, the cycle length to be applied to the sub-system during the subsequent cycle is given by

$$C'' = C + C' \qquad\qquad (2.29)$$

The function $f(C)$ is numerically equal to the value of DS which will give $C' = 0$, at the cycle length C currently operating, that is, the equilibrium DS at a given cycle length. The value of f(C) is low at low cycle lengths and high at high cycle lengths so that cycle length increments are more easily achieved at low cycle lengths.

In SCATS, for each subsystem, four cycle lengths are assigned: a minimum cycle length (Cmin), maximum cycle length (Cmax), a medium cycle length (CS) at which good two-way coordinations can be achieved in the subsystem, and a cycle length (CX) above which all additional cycle length is given to a nominated phase, known as the "stretch" phase. In cases where C > CX, only those detectors on approaches which will benefit from the increasing proportion of green are eligible to contribute a DS measurement for the calculation of cycle length. The system dynamically adjusts cycle time to maintain the highest degree of saturation in a coordinated group of signals within acceptable user defined limits.

2.6.1.2 Green Split Plans. Four green split plans are specified for each intersection identifying the proportion of the cycle to be allocated to each phase. The split plan also specifies the sequence of phases. The sequence of phase may vary between plans. A number of options can exist for each phase to control the transfer of unused time between phases and to ensure that nominated phases are included in or excluded from the sequence as a function of cycle length. The phase splits are modified by varying the cycle length. As the cycle length increases, the phase whose time is nominated in seconds receives a reduced proportion of the cycle. Phases nominated as a percentage of cycle lengths and the stretch phase receives a constant proportion of the cycle until the cycle

length reaches the value CX. Above CX, all additional cycle time is given to the nominated stretch phase.

The selection of the green split plan is done once per cycle by a split plan vote. The vote is obtained by calculating the value of DS that would have occurred on each approach if the new plan was operating. If two votes for the same plan in any three consecutive cycles are obtained, a particular plan is selected. The plan that yields the lowest value of DS receives the vote. The combined action of the cycle length and plan vote algorithms seeks to equalize the value of DS on all strategic or competing approaches.

2.6.1.3 Offset Plans. Internal offsets for each intersection are provided by a set of five plans, while the offsets between subsystems are defined in five external offset plans. An offset within SCATS, specified in seconds, is inherently independent of cycle length, but may be modified as a function of the cycle length to accommodate queuing or link speed changes during heavy traffic. The pattern of offsets in a series of coordinated signals must be varied with traffic demand to travel through a network of signals. SCATS selects offsets, based on free flow travel time and degree of saturation, which provide minimum stops for the predominant traffic flows.

2.6.2 Operation of SCATS

Control within SCATS is attained at two levels namely, "strategic" and "tactical". Together these levels of control determine the three principle timing parameters of traffic signal coordination: cycle time, phase split, and offset. Strategic control is concerned with the determination of suitable signal timings for the areas and sub-areas based on

average prevailing traffic conditions, while tactical control refers to control at individual intersection level within the constraints imposed by the regional strategic control.

2.6.2.1 Strategic Control. The phase splits and cycle time are calculated for the critical intersection and offsets are determined by the amount of traffic flowing in each direction through the sub-system. Phase splits for minor intersections in the sub-system are, by definition non critical, and are therefore either non variable or selected by a matching process which selects splits which are compatible with the splits in operation at the calculated intersection. Coordination over larger groups of signals is obtained by linking together sub-systems to form larger systems, operating on a common cycle time. The links determining the offsets between the sub-systems may be permanent or may link according to varying traffic conditions. This ensures that the traffic flow between sub-systems is sufficient to warrant coordination, the link is enforced, but if one or more subsystems can operate more efficiently at a lower cycle time, the link is broken.

From the DS measured for each lane of strategic detection, a normalized flow rate is calculated which is analogous to passenger car unit (PCU) flow. Cycle time is adjusted to maintain the degree of saturation around 0.9 on the lane with the greatest degree of saturation. Lower and upper limits of cycle time are specified by the user. Cycle time normally can vary by up to 6 seconds per cycle, but this limit increases to 9 seconds when fairly permanent trends are identified. Phase splits are varied by up to four percent of cycle time each cycle so as to maintain equal degrees of saturation on the competing approaches, thus minimizing delay. Best offsets are selected for the high flow movements. Other links carrying lower flows may not receive optimum coordination

when the cycle time is inappropriate. As and when traffic conditions permit, cycle times are maintained to values which provide good offsets on the majority of links.

2.6.2.2 Tactical Control.   Tactical control provides for green phases to be terminated early when the demand for the phase is less than the average demand and for phases to be omitted entirely from the sequence if there is no demand. Since all controllers in a system must share a common cycle time to give coordination, usually the main road phase cannot skip or terminate early. Any time saved during the cycle as a result of other phases being skipped or terminated early may be used by subsequent phases or is added on to the main phase to maintain each local controller at the system cycle length.

The United States Department of Transportation began the implementation of a national demonstration project for suburban ATMS utilizing SCATS in the late 1990s. The project was undertaken in the city of South Lyon, Michigan, which had an existing well timed fixed signal plan and the project involved conversion of all six signals to SCATS control. A research study *(Wolshon and Taylor, 1999)* was carried out to analyze the differences in certain delay parameters which would occur as a result of implementing SCATS signal control.

The methodology involved in this study tested the alternative traffic control strategies against identical traffic patterns. The comparison of total delay during the SCATS and simulated periods in this research were accomplished using a macroscopic simulation model. The model employed the delay calculation procedure found in the Signal Operations Analysis Package (SOAP) analysis program (USDOT, 1979). The research first analyzed traffic conditions under the control of SCATS system, and then used the same volume conditions and cycle lengths to determine the phase plan for a

fixed time signal system that would have minimized delay during the previous time period. The program SOAP was developed to analyze traffic signal alternatives at four-legged intersections with or without protected left turning phase interval in the signal sequence, including fixed time, semi-actuated, and fully actuated control. Thus, a macroscopic delay calculation procedure based on the SOAP procedures was used to determine the average approach delay and total intersection delay for each hour of the day.

Eight sample periods were selected to analyze the performance differences during peak, moderate and low traffic volume conditions. Results indicated that the average total intersection delay, on a system wide basis, was higher under SCATS control than under the simulated fixed time control strategy. The reason for higher delay is that the objective of SCATS is to equalize saturation flows rather than to minimize total intersection delay. The overall increase in the total delay was due to small increases in delay experienced on the major approaches especially during the peak traffic volume periods. The second conclusion indicated that SCATS tended to distribute the average approach delay more equitably to the constituent intersection approach movements. The third conclusion was that there was more effective reduction of delay during low volume periods compared to high volume periods under SCATS control.

Another critical evaluation of performance of SCATS against TRANSYT was carried out at Parramata, Australia in 1980, which covered 22 signals *(Luk, et al., 1982)*. The SCATS system comprised of a central monitoring computer, twelve remote regional computers and 600 microprocessor local controllers. After the collection of data for the traffic system under the control of SCATS and TRANSYT, the Statistical Package for the

Social Sciences (SPSS) was employed in the analysis of variance and covariance. The Student's t-test was used for comparing the means of journey times or stops for the two alternative strategies. The Student's t-test and parallel regression were used to compare the means of paired journey times or stops after analysis of covariance was used to correct for the effect of traffic flow. It was found that disregarding the error affecting the operation of SCATS in certain streets, SCATS successfully optimized journey times to a level consistent with that achieved by TRANSYT during the survey period and achieved a large reduction in stops. It was understood that since SCATS was required to adapt to the traffic conditions prevailing outside the survey hours, expected benefits could be achieved in a long term.

2.7 Optimization Policies for Adaptive Control (OPAC)

Optimization Policies for Adaptive Control (OPAC) is a computational strategy for achieving a demand-responsive traffic signal control in real time. It is a real-time signal timing optimization algorithm which was developed at the University of Massachusetts, Lowell (UML) in 1986. The strategy provides a dual capability to control distributed individual intersections as well as coordinated control of intersections in a network. The strategy serves as a building block for demand-responsive decentralized control in networks *(Gartner, 1986; Gartner et al., 2001)*. The OPAC strategy has developed in four versions.

The first version of OPAC, called OPAC-I served as the basis for the development of future versions of the strategy. This earliest version used a Dynamic Programming (DP) approach for solving traffic control problems. Dynamic Programming

is a mathematical tool to obtain optimization of multistage decision processes. In DP values affecting the process are optimized in stages rather than simultaneously. Although the approach provides global optimum solutions, it requires an advanced knowledge of arrival data over the entire control period. It is difficult to obtain this data with reasonable accuracy, and the optimization procedure requires an extensive computational effort since it is carried out backward in time. Thus, DP is not suitable for on-line applications.

The shortcomings of DP based optimization were overcome by OPAC-2, which consisted of a simplification of the OPAC-I algorithm. It served as a building block in the development of a distributed on-line strategy. In this version, the control period was divided into sequential stages of T seconds, where the stage length was equal to a typical cycle length, though it could be longer. The typical range was taken as 50-100 seconds. Each stage was divided into an equal number of intervals where interval length, s, varied from 2-5 seconds. The algorithm provides enough phases for each stage, so that no optimum solution is missed. The phase change times of time are measured from the start of the stage in time units of seconds. For any particular switching sequence at stage n, the performance function for each approach is the sum over all intervals in the stage of the initial queue length plus the arrivals minus departures during each interval. This is represented in Figure 2.5.

Thus, for any given switching sequence at stage n, a performance function is defined on each approach that calculates the total delay during the stage. Graphically, total delay is associated with the area enclosed between cumulative arrivals and cumulative departure curves, and represents the queuing of vehicles. The objective

function calculating the total delay is sequentially evaluated for all feasible signal change, or switching sequences.
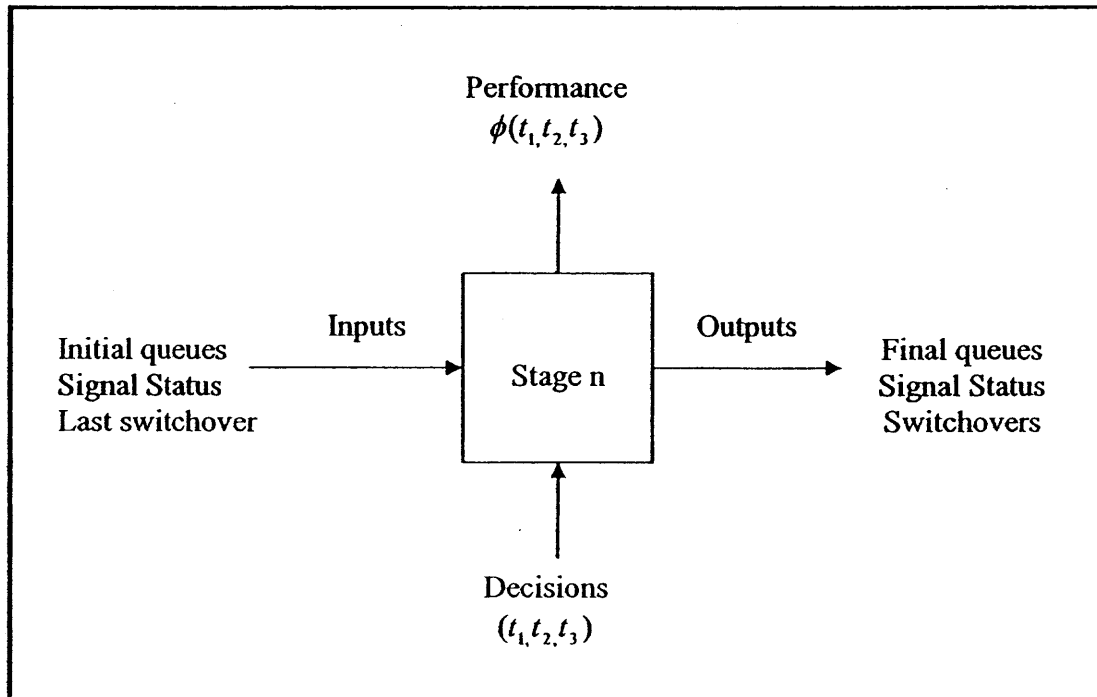
Figure 2.5 Information and decision flow at stage n of simplified optimization procedure.

The optimal value of total delay is obtained by a search algorithm, called Optimal Sequential Constrained Search (OSCS) method. The method searches for all possible combinations of valid switching timings within the stage to determine the optimum set. The method compares the current PI with the previously stored value of PI for all iterations. The system also stores the corresponding switching point times and final queue lengths. The optimal signal change or switching policies are calculated independently for each stage in a forward sequential fashion for the entire process.

The on-line application of OPAC-2 required knowledge of arrivals over the entire stage, i.e. a period of 1-2 minutes long. Thus it could be implemented with a traffic

prediction model which predicts the traffic arrival pattern over the entire stage. Since experiences have shown that prediction algorithms are some times less effective than the historical data, the concept of a rolling horizon was introduced in OPAC-3. The rolling horizon concept was introduced to reduce the unreliable nature of the future arrival information. In the rolling horizon technique, a stage length "k" and a roll period "r" are selected. Then, flow data is obtained for the first "r" intervals by detectors, and the flow data for "k-r" intervals is calculated from the model or detectors. An optimal switching policy is calculated for an entire stage, but is actually implemented only for the roll period. Finally, the projection horizon is shifted by "r" units to obtain a new stage, and the entire procedure is repeated, as shown in Figure 2.6. The rolling horizon technique shifts the horizon by certain units called rolled periods, after implementing the optimal policy for a particular stage. At the conclusion of any head period, a new projection horizon containing new head and tail periods with the beginning of the new horizon period at the termination of the old head period. The roll period can be equal to one or more number of steps. The calculations are repeated for each new horizon period. The actual arrival information over the head period can be obtained by placing the detectors well upstream of the intersection (10-15 sec travel time). Thus, OPAC relaxed the reliance on predicted or historical data. The actual arrival information allows a calculation for phase change decisions.
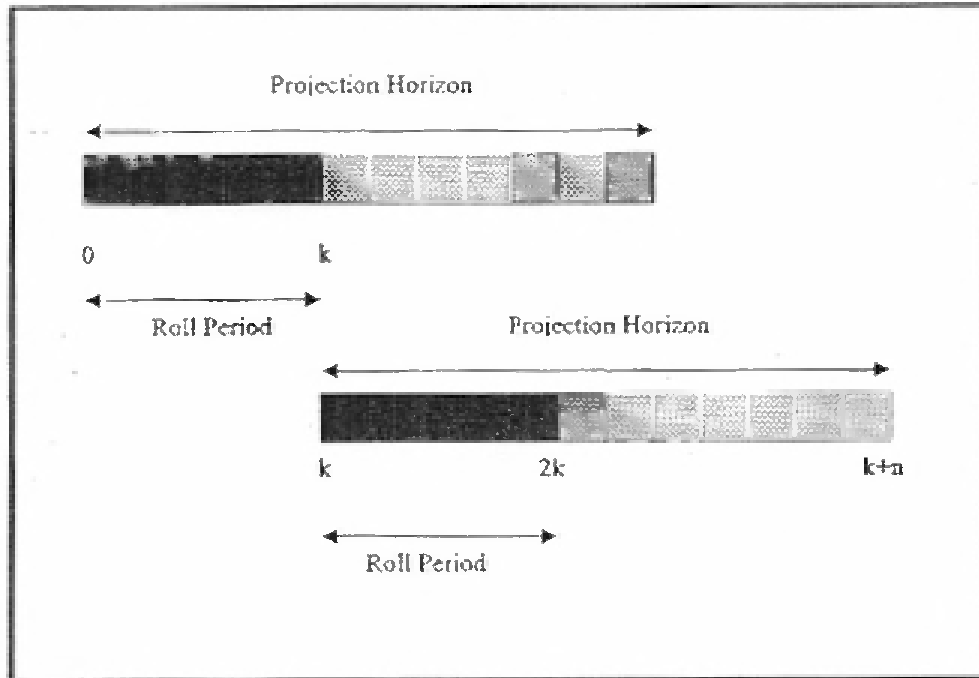
Figure 2.6 Implementation of the rolling horizon approach in OPAC.

Two types of models for the tail have been tested in OPAC (a) variable-tail, and (b) fixed tail. In the variable-tail, projected actual arrivals are taken for the tail, and in the fixed tail, the tail consists of a fixed flow equal to the average flow rate during the period. The first model was used to test the validity of a rolling horizon concept by comparing the results with previous experimentations. The model replicated the standards obtained with DP approach. The second model represented the practical approach to implementing OPAC. The model use upstream detector measurements for the head data and smoothed average flows for tail data. The head data is continuously updated in the rolling process. This model provided results close to the optimal and hence, represented a simplified and feasible approach to real-time control. Thus, it was concluded that OPAC offered substantial savings as compared to some of the prevailing fixed time strategies.

The OPAC project was expanded by UML by introducing a user defined option of coordination/synchronization strategy suitable for implementation on networks and arterials. This version of OPAC is referred to as Virtual Fixed Cycle-OPAC (VFC-OPAC), since the local cycle reference point, or yield point from cycle to cycle, is allowed to range close to the fixed yield point dictated by a virtual cycle length and offset. This enables the synchronization phases to better manage traffic conditions by an early termination or extension. This version of OPAC is composed of a three-layer control architecture. Layer 1 is the Local Control Layer and it implements the rolling horizon procedure and continuously calculates optimal switching sequences for the projection horizon, subject to constraints from Layer 3. Layer 2 is the Coordination Layer which optimizes once per cycle, the offsets at each intersection. Layer 3, the synchronization layer calculates the network-wide virtual-fixed-cycle once every few minutes, as specified by the user. The flexible cycle length and offsets are updated as the system adapts to changing traffic conditions. If desired, the cycle length can be calculated separately for groups of intersections.

A field test of OPAC was conducted using a special version of the NETSIM simulation model in Tucson, Arizona. Five different 30-min data sets of a signal-controlled intersection were tested. The OPAC policies proved a reduction of 30-50 percent of the initial delay. The improvements are noted in terms of corresponding speeds, which were averaged over all links of the simulated network. The test included large portions of travel time that were not subject to influence by the control strategy. The increase in average speeds ranged from 10 to 20 percent.

In 1992, the FHWA started to develop a state-of-the-art system, Real-Time traffic Adaptive Control System (RT-TRACS) in conjunction with other agencies *(Gartner, et al., 2001)*. The objective of RT-TRACS was the development of a system capable of adapting to fluctuating traffic conditions as they occur by selecting the optimal control strategy from a "suite" of real-time traffic signal timing control strategies. In 1998, the first version of RT-TRACS using OPAC was implemented on a network of 16 intersections on Reston Parkway in Reston, Virginia. The Reston RT-TRACS system controlled this network by a distributed system of 2070 controllers. The preliminary findings indicated that OPAC has achieved the objectives of reducing stops and delays with a considerable degree of success and provided progression along arterial.

2.8 Summary

This chapter gave an overview of the history and development of adaptive control in traffic engineering and a detailed review of some of the existing strategies. It gives a classification of the strategies, related constraints, requirements for the optimum performance of these strategies and their shortcomings. The chapter also gave a detailed description of the ongoing use of control theory to manage congestion within arterials, freeways, ramp metering and also as an optimization technique for route guidance purposes. A detailed review of the function and operations of three major urban adaptive control strategies, SCOOT, SCATS and OPAC was also included. The literature search indicated that most of current adaptive control strategies which work on the feed-forward principles might enhance their performance by the introduction of concepts of feedback control systems. The feedback property of control theory can reasonably address the

problems of existing adaptive control strategies like SCOOT and SCATS. This chapter contributed to the dissertation by giving a fair idea of the major short comings of existing strategies, requirements of the current traffic conditions, and the effectiveness of control theory and controllers for congestion management. It also helped to develop ideas for the choice of control variables and an appropriate methodology for the mathematical formulation presented in Chapter 4.

CHAPTER 3

METHODOLOGY

This chapter provides a detailed description of the basic framework for introducing the concept of feedback control into adaptive traffic control systems. The conceptual approach is supported by describing an application of the concept to the existing adaptive traffic signal strategies: SCOOT, SCATS and OPAC. The intent is to demonstrate that by incorporating the concept of feedback control to the basic operation of these systems, their performance is enhanced. This discussion is followed by a detailed review of the concept of feedback control, its components, types and operation of feedback controllers.

3.1 Introduction

This research uses feedback control theory to enhance the performance of adaptive traffic control systems. The incorporation of principles of a feedback control can overcome some of the shortcomings associated with the inaccurate estimation and evaluation of flows in traffic models used by the existing adaptive control systems.

3.2 Elements of an Adaptive Control System

A sequence of steps is undertaken by most adaptive traffic control systems to complete the process of predicting the on-site traffic conditions and provide the best online signal timing plan. The elements and the flow of data in a general adaptive control system are shown in Figure 3.1. Adaptive control begins with the collection of a variety of data such as volume, speed and approximate queue lengths to determine traffic demand. The data

are used to determine the most appropriate control plan for the area under study. Once the demand is determined, the decision making technique selects the appropriate signal timing parameters to achieve a desired objective. The desired objective, depending on the type of algorithm and the demand condition, varies from minimizing delay and queues to maximizing throughput. All estimation and prediction of variable quantities, as well as procedures for optimization and selection of the best plan takes place in this step.
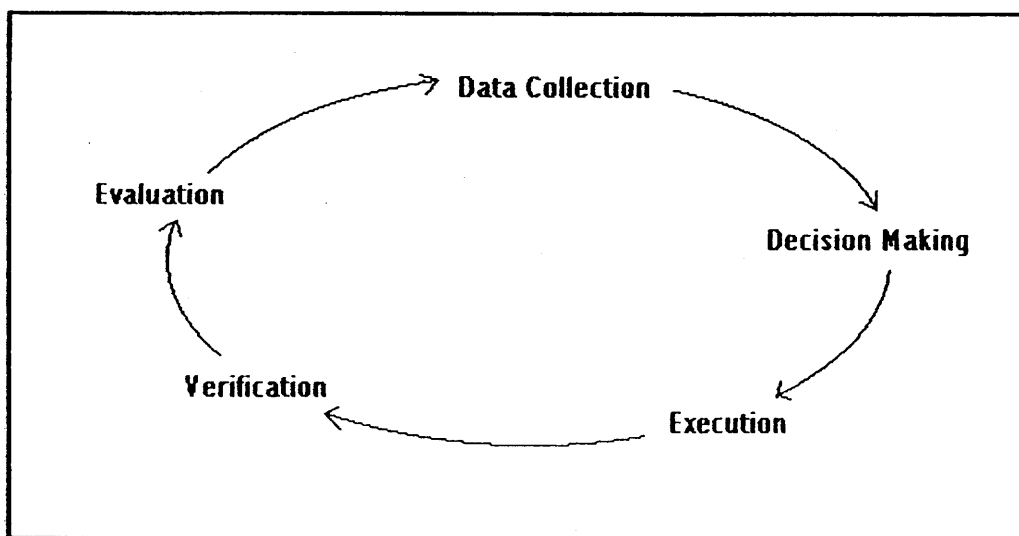
Figure 3.1 Elements of an adaptive control system.

The algorithm implemented at this stage typically calculates a detailed signal control plan for the study area. This is followed by the execution of the timing plan by downloading the signal timing plan to the signal controller. In the next step of the cycle, verification is performed by the control system to ensure that the calculated signal timing plan has been executed correctly by the controller. Verification is followed by the evaluation of the implemented plan to determine if the objectives of the control algorithm were achieved. This is typically carried out by comparing the results from implementing the new control plan with the results from the existing plan.

The five stage process described above represents the sequence of activities that need to be performed by a traffic control system to obtain adaptive control at signalized intersections. The sequence of steps described above represent a feed-forward manner of data flow within the adaptive traffic control system because the result of an iteration is not fed back in to the next iteration, and thus, the results of two iterations are independent of each other. However, in real world traffic flow conditions, traffic flow pattern in one cycle length may affect the flow pattern of subsequent cycles. This approach lacks an efficient mechanism to update estimated quantities in case of unpredictable conditions. This data flow fails to efficiently implement the evaluation stage of the procedure of adaptive traffic control because the feed-forward data flow system fails to consider the effect of factors such as residual queues during the estimates involved in subsequent time periods. This research improves the performance of current adaptive control systems by introducing a feedback control loop to overcome the deficiencies of feed-forward data flow within these systems, as shown in Figure 3.2. Hence, the decision making phase of the existing adaptive control system is enhanced by adding a feedback loop.
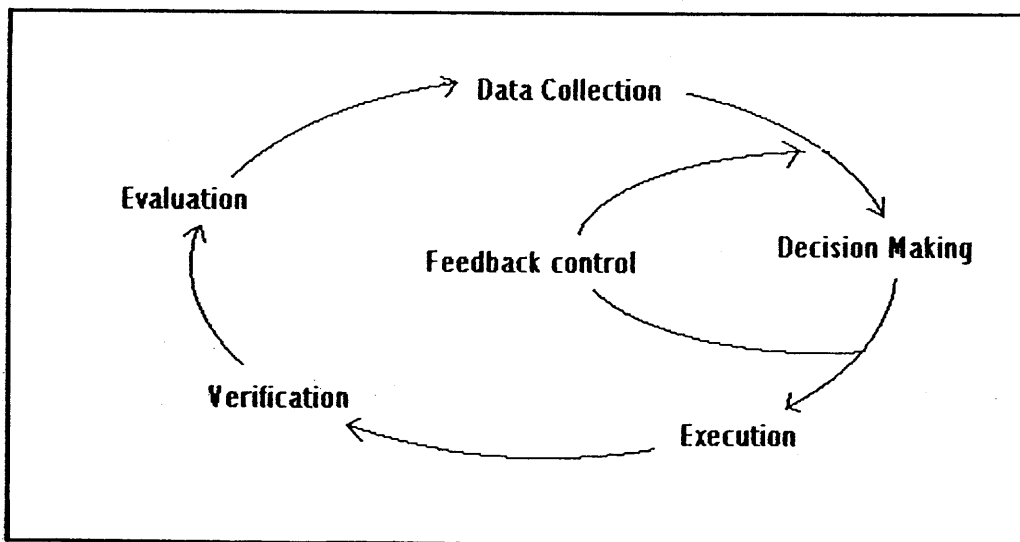
Figure 3.2 Proposed enhancement to the existing adaptive control system framework.

3.3 Problem Description in SCOOT

The existing adaptive traffic control strategy considered for applying feedback control is SCOOT. The system uses the concept of Cyclic Flow Profiles (CFPs), which are measured as they occur on the street rather than calculated offline. The concept of CFPs in SCOOT is borrowed from the TRANSYT model. Cyclic flow profiles represent the measure of the average one-way flow of vehicles past any chosen point on the road during each part of the cycle time of the upstream signal. The CFPs are used to estimate the number of vehicles that will reach downstream signals, as well as to estimate the queue lengths and time required to clear the queues. Hence, CFPs play a key role in delay estimation within these models.

The simulation of junctions in SCOOT occurs in an upstream to downstream manner. This technique of simulating traffic conditions within a network works efficiently during undersaturated conditions. However, during peak demand conditions, as the queues start building on the links and spillback conditions start appearing, the cyclic flow profiles estimated on the basis of traffic flow from upstream to downstream tend to become less accurate. This is caused due to the inaccurate prediction of arrival flow at a stop line due to saturated conditions at the downstream links. While simulating a particular link, the SCOOT simulation model does account for the links that are saturated with existing queues. TRANSYT also accounts for saturated links and denotes them as "full" links with no inflow allowed into them. Similar to TRANSYT, SCOOT also accounts for the outflow pattern of the upstream links to estimate the inflow pattern of subject links. During a continuous flow of traffic, the SCOOT model does not

adequately account for the effect of over saturation in the downstream link on the arrival flow from the upstream link.

This flow of data from the upstream to downstream intersection eliminates the possibility of an interference of the downstream link vehicles with the vehicles in upstream links. However, this approach of simulation does not seem to adequately evaluate the traffic flow patterns during oversaturated conditions as the outflow pattern of the downstream link plays an important role in the arrival flow pattern of the upstream link. The outflow pattern of the downstream link becomes vital for the oversaturated conditions due to the presence of queues partially or fully occupying the downstream links and prohibiting the estimated flow rates to leave the upstream links.

The arrival flow pattern within TRANSYT is expressed mathematically as follows:

$$IN_{it} = \sum_{j}^{n} F_{ij} \left(P_{ij} \bullet OUT_{jt'} \right) \tag{3.1}$$

where, IN_{it} represents the IN-pattern on link i for time step t, F_{ij} represents a smoothing process related to platoon dispersion for flow to link i from link j, P_{ij} describes the proportion of the feeding link OUT-pattern that feeds the subject link, $OUT_{jt'}$ indicates the OUT-pattern of link j for step t', t' is the time step t minus the travel time to link i from link j and, n the number of links (j) that feed link i.

Since the expression for the arrival flow pattern does not consider the out pattern of the downstream link, the arrival flow pattern of the upstream links becomes inaccurate and the inaccuracy is carried backward to the upstream links as simulation progresses to the next time period. The inaccurate arrival flow pattern, as described above, requires an update to adjust the CFPs for the following time step. Version 8.1 of TRANSYT

simulates queue spillbacks and queue blockades and includes the associated penalties in the optimization delay functions. The procedure, however, does not adequately provide a methodology to control the buildup of spillback and queue blockade conditions.

This shortcoming of the CFPs can be overcome by a properly designed feedback control system. The arrival flow patterns can be adjusted upstream by feeding the output of the downstream link back to the arrival pattern from the upstream. In other words, incorporation of a feedback control system can facilitate the adjustment of cyclic flow profiles by adequately accounting for the updates required in the CFPs on upstream links during the onset of oversaturation at the downstream. This can be done by adding the outflow pattern from the downstream into Equation (3.1). The output from the feedback control system can be used to obtain the outflow pattern downstream. A feedback control system can also serve as a mechanism to reduce the possibility of occurrence of spillback and queue blockades. By constantly updating the flow profiles and thus the arrival patterns, the controllers of a feedback control system can suggest small changes in the optimized signal plan provided by the TRANSYT optimizers.

The following sections provide an introduction to the basics of feedback control systems, describe various types of controllers and present a brief overview of z-transforms which are an important tool for the mathematical representation of feedback control systems.

3.4 Introduction to Feedback Control

Control systems are generally divided into two categories: open loop and closed loop. An example of an open loop system is a fixed time signal which goes through the same set of predetermined green-amber-red signal sequence irrespective of the demand raised at the main or cross streets. A signal control system operated by a policeman is a closed loop system in which the change of signals is done according to the demand raised at the cross street. Hence, a closed loop system operates with a certain desired output and tries to control the input and other variables in such a way that the desired output is achieved. Closed loop control systems have been known to be most robust and stable to changes external to the system because of their feedback nature *(Belanger, 1995)*.

Control systems can also be classified as continuous or discrete. In discrete-time systems, or sampled-data systems, one or more variables can change at discrete instants of time. Continuous-time systems differ from discrete systems in that the signals for the latter are in discrete sampled data form, i.e. the signal is discretized in equal time steps, while the former uses continuous signals. Here, a signal refers to the input or output from the modules used in the block diagram of a feedback control system. A signal graphically represents the pattern of the input or output.

The main elements of a control system include input variables, the outputs, and controllers used to carry out the control strategy. The control system may include human monitoring and intervention, which is achieved through an operator interface. The following section gives a brief overview of the basic components of a feedback control system and their function with respect to the traffic signal system under study.

3.5 Basic Components

The elements of the control system correspond to some physical process. In terms of a traffic signal system for an arterial network, the input signal is the input volume of vehicles into the arterial. The sensors of the control system correspond to vehicle detectors which measure input volumes. The output is volume discharged by the system during a time index k. The signal system, which controls the movements of vehicles at an intersection, corresponds to the plant of the control system. Figure 3.3 illustrates the basic components and the flow of data within a control structure *(Belanger, 1995)*.

Inputs: The inputs in a control system can be manipulated or controlled, thus when used in a control plant, they are called controllable inputs or simply inputs. The input in a closed-loop system is fed into a mixing point along with a feedback signal coming back from the output, which is the controlled quantity. The input in the system under study is the vehicular volume to be released by a controller from the upstream intersection into the downstream link.

Disturbance: The inputs, which cannot be manipulated or controlled, are called disturbances. These disturbances influence the plant output, thus it is desirable to design a control system such that the disturbances have a minimal effect on the system. Disturbances can be predictable or unpredictable, measurable or nonmeasurable quantities. The uncontrollable disturbance in our system is the volume of upstream through vehicles and turning vehicles on to the downstream intersection.
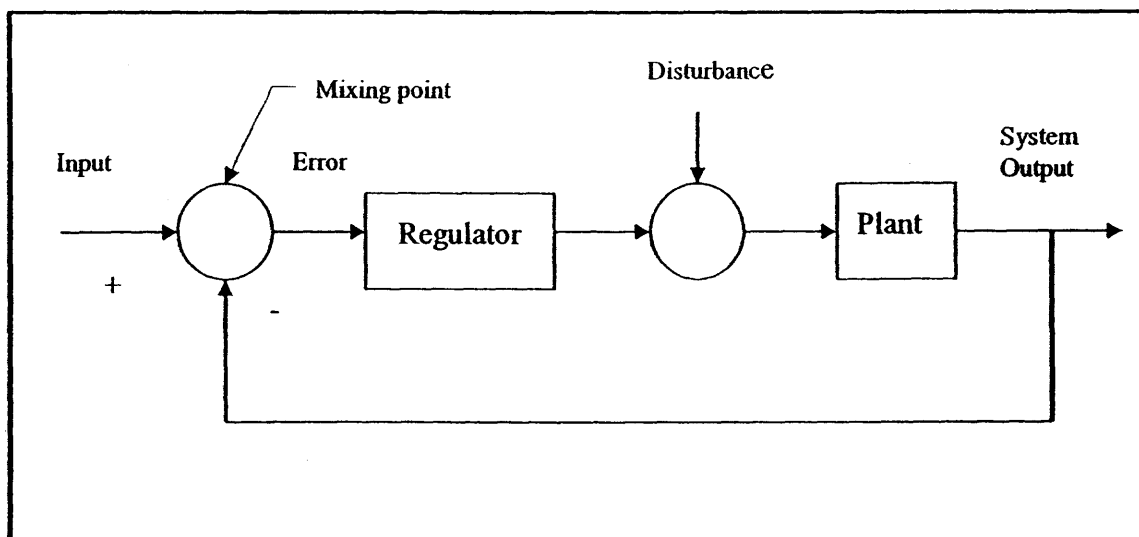
Figure 3.3 Closed loop feedback control.

Mixing Point: The mixing point calculates the difference between two control signals having the same units. This difference is referred to the error in the system. The error acts as the trigger for the functioning of the control system. Thus, the mixing point has the input and output signals of the control system as its inputs. The signal which comes from the mixing point is also called the actuating signal. According to the concept of a closed loop system, the controlled quantity is affected by the disturbance. The actuating signal is affected by the output and tries to correct the controlled quantity.

Error: In an ideal condition, it is desired to have a controlled quantity exactly follow the referenced input. But if the output does not take on the desired value, it has an error, which, as previously discussed, is equal to the difference between its desired and actual value. In the case of a zero error, the closed loop system is said to have reached a steady state. The error term in the system under study is the difference between the capacity at an intersection (desired value), and the actual output. It is desired to operate intersections at capacity during oversaturation conditions. However, a variety of factors contribute to a

throughput less than the capacity. Traffic signal systems operating in a feedback environment can be timed to allow the discharge volumes from intersections in the system to track the capacity of the intersections with the help of feedback control. The difference between the desired and the actual output, or the error of the system, acts as an actuator for the feedback control loop to function and track the desired capacity condition.

Regulator: Controllers or regulators are used to make the system track a certain desired trajectory. A controller, in the form of an algorithm, handles the disturbances and brings the system to a desired steady state. Although, it is practically impossible to completely eliminate the steady state error, the controllers try to stabilize the system up to a reasonably stable state within some transient time by controlling the input of the system based on the desired performance and outputs. The steady-state error is defined as the limit of the difference between the input and output of a system as time goes to infinity (i.e. when the response has reached the steady state) and can be expressed as:

$$error = \lim_{t \to \infty}(input - output)$$

The steady-state error largely depends on the type of input (step, ramp, etc). The function of a controller is to minimize the steady state error. The steady state error for the system under study is the difference between the discharge capacity of one of the links at the downstream intersection and the actual output from the intersection in that link.

3.6 Z – Transforms

Classical control theory extensively utilizes transfer function and frequency domain concepts for design and analysis, while modern control systems use the state-space concept based on vector-matrix analysis. The design and analysis of modern control systems is done in a time domain. For this study, the classical transfer function technique was used for the design of the proposed control system. Depending upon the nature of the system, i.e. continuous or discrete, Laplace or z-transformation techniques are used respectively for the design. Laplace transform is used to find the solution of linear differential equations while z-transforms are used to find the solutions of linear difference equations. A transfer function is an operational method of expressing the differential or difference equations that relate the output variable to the input variable. Transformation techniques are used in this study since it is proved to be a time saver for the solution of linear differential/difference equations *(Muth, 1977)*. Another reason for the use of transformation techniques is that MATLAB is an efficient tool to simulate the control system and has the ability to include both Laplace and z-transformation techniques in the development and simulation of control systems. Use of transformation techniques to solve discrete and continuous time systems is well supported in MATLAB.

The flow of a signal between various components of the control system is expressed in the form of differential equations or difference equations depending on whether a continuous or discrete time system is modeled. Differential equations are functional equations that define sequences, while difference equations, which are a discrete counterpart of differential equations, define functions. The Z-transform is an operator that maps a sequence f_n, an element of the original space, into the function F(z),

an element of the transform space. The interrelation between the original space and transform space is shown in Figure 3.4. Thus the function F(z) is called the transform or image of the sequence f_n.

The transformation is expressed as:

$$Z\,[f_n] = F(z) \qquad\qquad (3.2)$$

The inverse transformation is written as:

$$Z^{-1}[F(z)] = f_n \qquad\qquad (3.3)$$

Both the Z-transform and its inverse are unique, so they can be expressed as the two-way transformation in double arrow notation as follows:

$$f_n \leftrightarrow F(z)$$
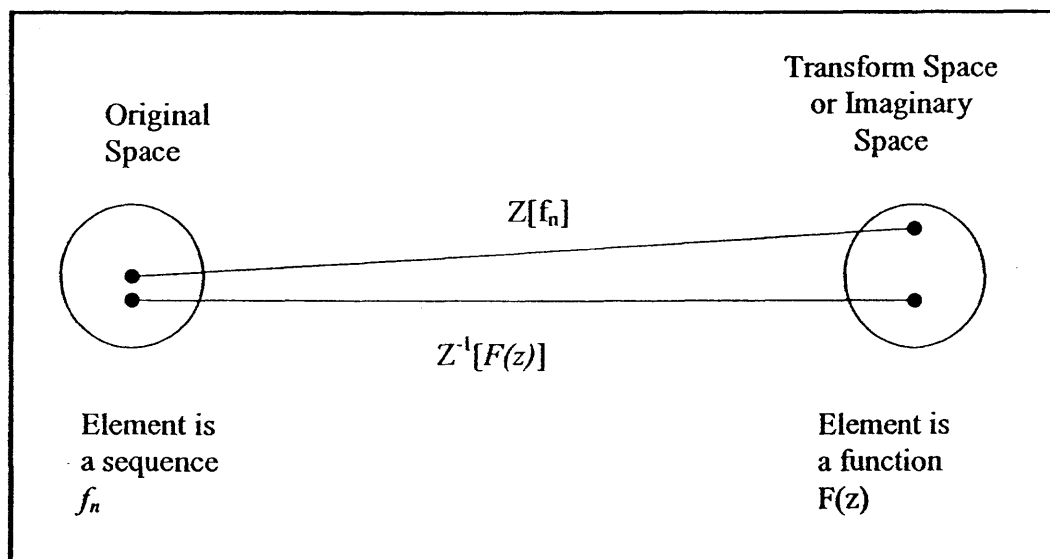
Figure 3.4 Transformations.

3.7 Classification of Controllers

There are several types of feedback controllers. Each controller has properties that make it suitable for certain conditions. These controllers when applied to a control system use their respective properties to bring the control system to steady state. A constant-gain controller, which is known as proportional controller, tends to reduce the error in the system, but worsens the steady state error properties of the system with increase in gain. An integral controller, when combined with proportional control, improves the steady-state error properties. Finally a derivative control has the property to improve the transient properties of the system. These three kinds of controls combined together are known as proportional-integral-derivative (PID) control. The PID control is a heuristic approach to controller design and is widely accepted in the process industries.
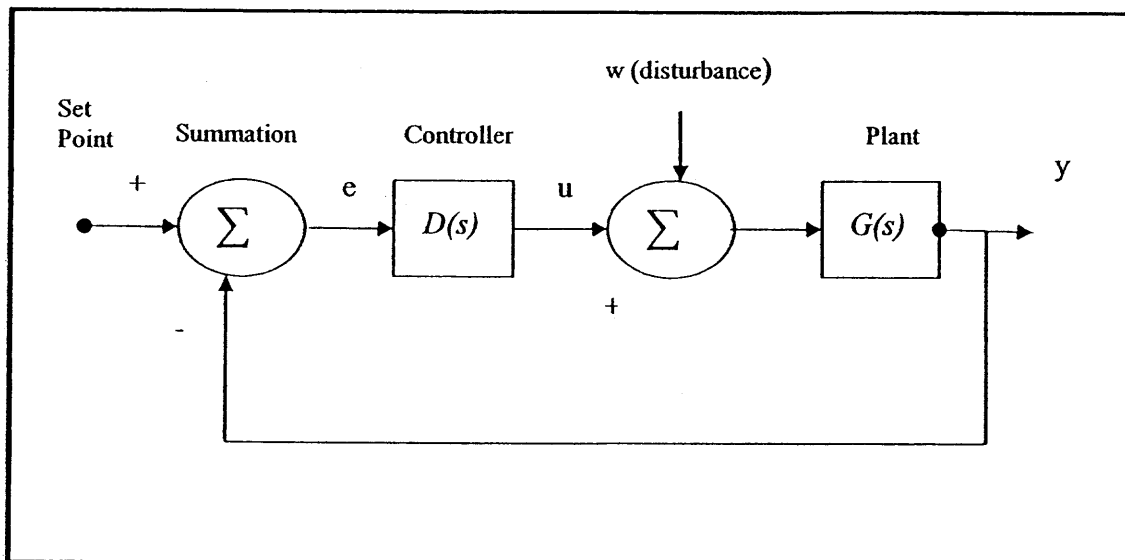
Figure 3.5 Feedback control block diagram.

The three types of controllers discussed in the following sections can be explained using Figure 3.5 where w represents the disturbance, e represents the error, y is the response of the disturbances and u, which represents the output from the regulator, is a

function of error e and the function varies depending on the type of control being applied *(Gene, 1994)*. In the figure, the controller has the transfer function $D(s)$ and the plant has the transfer function $G(s)$. A discussion on the transfer function applied in this study was provided in section 3.6. A description and mathematical representation of the types of controllers is presented in the following sections.

3.7.1 Proportional (P) Feedback Control

When the feedback control signal is linearly proportional to the error in the measured output, the result is called proportional feedback. The general form of a proportional controller is

$$u = Ke \tag{3.4}$$

where, K is the proportional gain, and e represents the error term. In this case, u is a function of the current value of error e. Therefore, the controller transfer function $D(s)$ that establishes a relationship between the input e and output u of the controller is given by

$$D(s) = K \tag{3.5}$$

Thus, a proportional controller is an amplifier with a gain that can be adjusted by varying the value of K.

3.7.2 Integral (I) Feedback Control

In a control system with integral action, the controller output is proportional to the amount of time the error is present. The primary function of an integral controller is to reduce or eliminate constant steady-state errors. The use of the integral controller comes at the cost of a worse transient response, i.e. the transient response characteristics like

delay time, rise time, peak time, overshoot or settling time, do not follow the specifications satisfactorily. The general form of an integral controller is

$$u(t) = \frac{K}{T_I} \int\limits_{t_o}^{t} e\, d\eta;$$ (3.6)

where, T_I is the reset time or the time for the integrator output to reach $1 \mathrm{x} K$ with an input of unity. Hence, $1/T_I$ is referred to as the reset rate. For a I control, u is a function of all past values of e rather than just the current value. The error "e" is integrated over time step η from time limit t_o to t.

Therefore, the transfer function D(s) of the controller from Figure 3.6 becomes

$$\frac{U(s)}{E(s)} = D(s) = \frac{K}{T_I s}$$ (3.7)

where s is a complex variable that has a real component σ and an imaginary component $j\omega$ and is represented as

$$s = \sigma + j\omega$$ (3.8)

3.7.3 Derivative (D) Feedback Control

Derivative feedback has the following form:

$$u(t) = KT_D\, \dot{e}$$ (3.9)

where, T_D is the derivative time and $\dot{e}$ represents the derivative of the error e. Derivative control is used in conjunction with a proportional and/or integral feedback controller to increase the damping and generally improve the stability of a system.

Therefore, the $D(s)$ becomes:

$$D(s) = KT_D s$$ (3.10)

3.7.4 Proportional-Integral (PI) Feedback Control

The PI controllers work by summing the current controller error and the integral of all previous errors. Equation (3.11) describes a PI controller output as

$$u(t) = K\left(e + \frac{1}{T_I}\int_{t_0}^{t} e \, d\eta\right)$$ (3.11)

3.7.5 Proportional-Integral-Derivative (PID) Control

For a PID controller, the control signal is a linear combination of the error, the time integral of the error and the time rate of change of the error. All the gain constants are adjustable. The controller transfer function is given by:

$$D(s) = K\left(1 + \frac{1}{T_I s} + T_D s\right)$$ (3.12)

The terms in Equation (3.12) express quantities that have been explained earlier. To design a particular control loop, the engineer has to adjust the constants K, T_I and T_D to arrive at acceptable performance. The adjustment process is called "tuning of the controller".

All the controllers discussed above have different characteristic properties called transient properties. These properties make a designer decide the type of controller most suited to a certain condition. The proportional feedback control can reduce error responses to disturbances by implementing a linear gain to the error in order to track the desired trajectory. However, it still allows a non-zero steady-state error. In addition, proportional feedback increases the speed of the response, but has a much larger transient overshoot. Overshoot is the difference between the peak value and the steady state value

of the control variable within a system. Definitions of the transient properties of controllers are provided in Appendix A.

When this controller includes a term proportional to the integral of the error, then the steady-state error can be eliminated, although this leads to further deterioration in the dynamic response. Dynamic response is defined as the direct or indirect effect on a complex system under the influence of a force, at the time and place of application of force or even otherwise. Finally, addition of a term proportional to the derivative of the error can dampen the dynamic response of a controller. Additional information on the properties of P, I and D controllers are provided in Appendix B. Appendix C provides an example of the application of the P, I and D controls to stabilize a simple mass-spring-damper problem.

3.8 Integrating Feedback Control in SCOOT

A feedback control system can be integrated into an adaptive control system as a means of evaluating the performance of the adaptive control system by feeding back the effects of one time period to subsequent time periods. Thus, the functioning of an adaptive control system in a feedback control environment can help in providing adequate and timely updates of control elements within the traffic signal, so as to improve the performance during subsequent time periods.

The SCOOT urban traffic control system in its current state is an example of a feed-forward control system. To accommodate the dynamism of traffic conditions at a signalized arterial, it is essential to employ more than one means of simulating traffic flow conditions, as well as predicting and mitigating the occurrence of spillback

conditions. Figure 3.6 is a representation of the proposed conceptual framework for using feedback control in an adaptive control system environment. The framework demonstrates the flow of information currently used in a feed-forward manner within SCOOT, but with an enhancement that adds a feedback control system to the existing modules of an adaptive control system. Within SCOOT, the enhancement targets the need to update the cyclic flow profiles of upstream junctions as a function of conditions at downstream links. The vehicle data is collected with the help of vehicle detectors in the form of volume per unit time. Based on this data, cyclic flow profiles are obtained. The cyclic flow profiles are then used by the traffic flow model to simulate the flow of vehicles within a link subject to platoon dispersion. Finally, the number of vehicles arriving at the downstream intersection and their arrival pattern is estimated. This information is used by the optimizers to obtain an optimized signal timing plan to serve the predicted volumes. Upon integration of this system with feedback control, the estimation procedure is improved. The vehicle flow data predicted at the downstream intersection is verified for existence of queues and if queuing exists, the feedback control system feeds back the queue information to the stage at which CFPs are calculated, so that demand can be accordingly modified for the subsequent time period. Figure 3.7 is a representation of details of a feedback control system within SCOOT in the form of a block diagram. The operation of this feedback control system can be explained with the help of this figure.
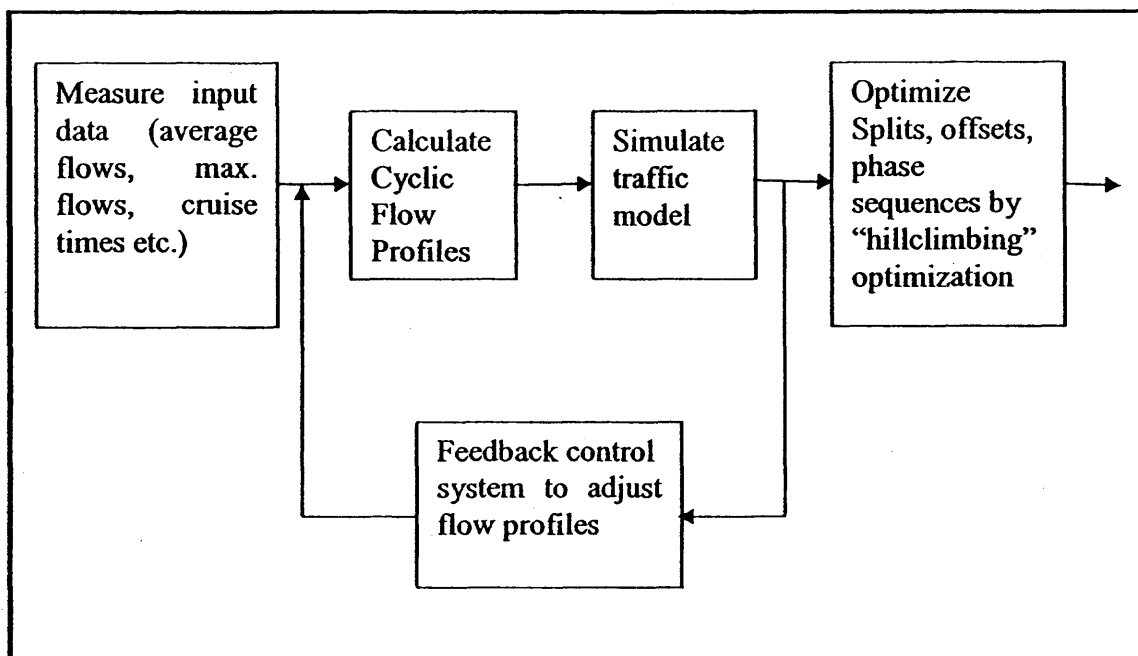
Figure 3.6 Conceptual framework representing an SCOOT in a feedback environment.
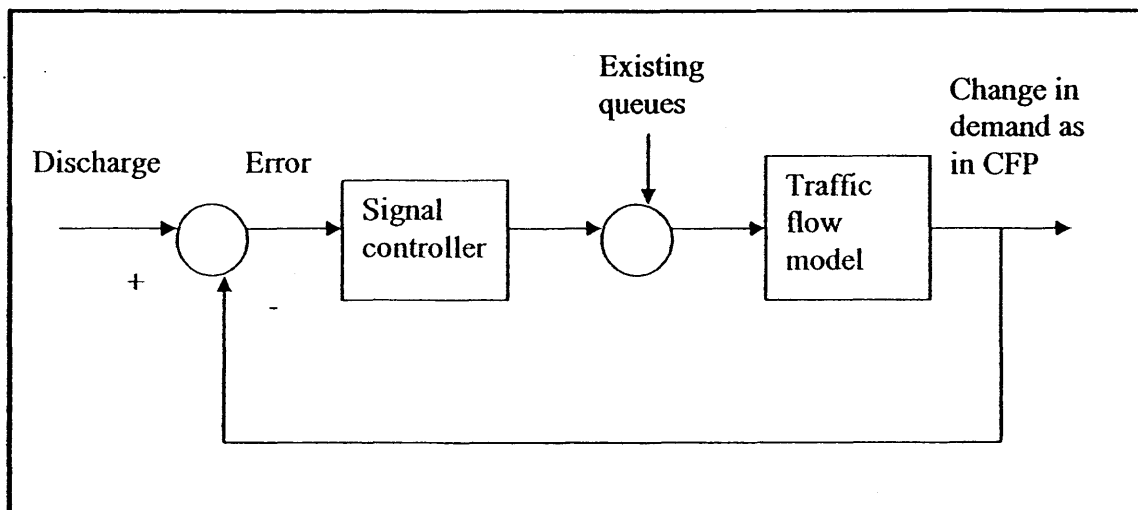
Figure 3.7 Details of feedback control system within SCOOT.

For a stepwise simulation, once per step, the number of vehicles discharged at the downstream intersection is fed into the feedback control system. As described in Figure 3.7, the reference input to the feedback system is the number of discharged vehicles from the downstream intersection. This input is a variable quantity and is directly obtained

from the detectors. The control system compares the reference quantity with the arrival flow rate in the upstream links. Depending on whether the difference is positive or negative, the feedback controller decides to make updates in the cyclic flow profiles accordingly. If the difference is positive, then the controller allows higher input into the upstream, if the difference is negative, the controller reduces the input into the upstream. This update helps to adjust the arrivals at the upstream in accordance with the capacity available at the downstream to accommodate these arrivals. The updated CFP represent the actual number of arriving vehicles. The time required to discharge· these vehicles moving under a particular saturation flow rate and density condition, can be calculated. Thus, the signal controller can adjust the signal timing plan in accordance with the changed cyclic flow profiles.

3.9 Problem Statement in SCATS

Another application of where feedback control can be integrated in adaptive traffic control system is in SCATS. The selection of cycle length in SCATS is dependent on the SCATS-defined term degree of saturation (DS). The DS is defined as the ratio of effectively used green to total provided green time. For application purposes, these green times are converted to the number of Passenger Car Units (PCUs) associated with the effective and available green times, taking into consideration factors such as speed and headways. The link on the network showing the highest DS contributes to deciding the cycle length for the next time period for the entire subsystem. Intersections within a particular subsystem are operational on this highest DS with minor adjustments of $\pm$ 6 seconds as per the specific requirement of each intersection.

As congestion increases during peak hours, the DS and cycle length continues to increase. There is no mechanism, however, to verify at the end of that cycle if the number of vehicles arriving at the intersection were able to discharge from the intersection within the cycle length. A variety of factors, such as mid-block sources and parking maneuvers can disturb the platoon, providing uneven flow of vehicles between intersections. Since the DS is based on the effectively used green time, the uneven flow of vehicles may lead to an inaccurate calculation of the effective green time. This leads to a DS, which does not correctly represent the actual demand at the upstream intersection. If the intended number of vehicles that should be discharged in a given cycle length is unable to leave the intersection and there is a residual queue, this may lead to inaccuracies in estimation based on the degree of saturation for the next time period, leading to upward traveling congestion. Over time, this condition can spread to cause spillback conditions on the upstream.

Feedback control can act as a mechanism to verify the actual number of vehicles discharged from an intersection, and to provide suitable corrections when the demand exceeds capacity and all vehicles are not able to be discharged in one cycle. Feedback control can then be used to determine the demand for the next time period. The geometric conditions of intersection or roadway network determines the queue holding capacity, and a prescribed cycle length can be associated with the number of vehicles that can be discharged within that cycle length. Depending on the DS, the intended number of vehicles that can physically leave the intersection in the given cycle length can be found. Including a feedback control loop within the basic framework of SCATS can be used to evaluate whether the intended number of vehicles were able to be discharged from the

intersection in the prescribed cycle length. If queuing exists, an estimate of the queue length can be made by calculating the difference between the actual number of vehicles corresponding to the previous and current cycle times. Provisions for accommodating the estimated queue length can be made in the next time period, by making adjustments to the prescribed cycle times.

Figure 3.8 shows the incorporation of a feedback control loop in the basic framework of SCATS. The control process begins with the first module that represents measurement of vehicle data by detectors. Once the desired data is obtained, the second module calculates the SCATS defined degree of saturation (DS). Based on this DS, the third module selects the appropriate cycle length, green splits and offsets from a pre-determined set of these traffic control parameters. Once the intersection operates according to these control parameters, queue length can be calculated based on the difference between the DS before and after the implementation of the control parameters. The residual queue is fed back by the feedback control system to modify the DS for the next time period. This is required because the measure of DS is used by SCATS to estimate the demand for the next time period. By considering the residual queues in addition to the new demand, the feedback control system enables SCATS controllers to better estimate the demand for the subsequent time periods during oversaturated conditions.
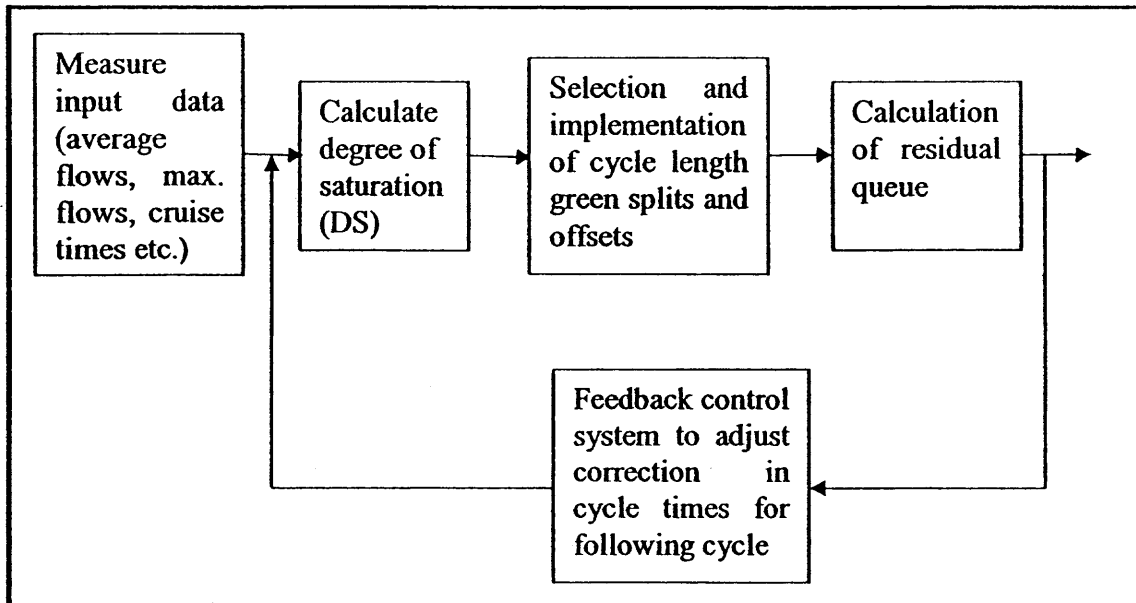
```
┌─────────────────────────────────────────────────────────────────────────┐
│  ┌──────────────┐   ┌──────────────┐  ┌──────────────┐   ┌──────────────┐ │
│  │ Measure      │   │ Calculate    │  │ Selection and│   │ Calculation  │ │
│  │ input   data │   │ degree of    │  │ implementation│  │ of residual  │ │
│  │ (average     │   │ saturation   │  │ of cycle length│ │ queue        │ │
│  │ flows,  max. │   │ (DS)         │  │ green splits and│ │              │ │
│  │ flows, cruise│   │              │  │ offsets      │   │              │ │
│  │ times etc.)  │   │              │  │              │   │              │ │
│  └──────────────┘   └──────────────┘  └──────────────┘   └──────────────┘ │
│                         ┌──────────────────┐                              │
│                         │ Feedback control │                              │
│                         │ system to adjust │                              │
│                         │ correction    in │                              │
│                         │ cycle times  for │                              │
│                         │ following cycle  │                              │
│                         └──────────────────┘                              │
└─────────────────────────────────────────────────────────────────────────┘
```

Figure 3.8 Conceptual framework representing SCATS in a feedback environment.

3.10 Integrating Feedback Control in SCATS

As explained, SCATS always tracks the demand denoted by DS in the time period succeeding the time period in which the demand occurs. Thus, a lag exists between the time at which demand is measured and when the signal plan is implemented to meet the demand. This leads to a likelihood that a queue will occur. To correct this condition, the property of feedback control can be integrated with the operation of SCATS during oversaturated traffic conditions. Whenever the degree of saturation is calculated as greater than 1, the SCATS algorithm calculates the queue length for the following time period in terms of PCU. This queue length can be used in a feedback loop to combine it with the number of vehicles corresponding to the DS of the preceding intersection for the previous time period. The combined value can give a more accurate measure of demand and the selection of the cycle length in the following time period. This approach can be explained as follows:

$$V'_{PCU}(k+1) = Q_L(k) + V_{PCU}(k+1) \tag{3.13}$$

where $V'_{PCU}(k+1)$ represents the total demand for time period $k+1$, $Q_L(k)$ is the residual queue from previous time period k and $V_{PCU}(k+1)$ is the demand for time period $k+1$ based on the DS for time period k. The demand $V_{PCU}(k+1)$ can be expressed as a function of the DS for time period k as follows:

$$V_{PCU}(k+1) = f[DS(k)] \tag{3.14}$$

This approach ensures a reduction in the possibility of queues. This approach also ensures an early clearance of queues and ascertains that queue buildup is not continued.

Figure 3.9 represents the block diagram of the feedback control loop proposed to be integrated into the SCATS framework. The DS obtained from the detector data is used to calculate the number of vehicles that can pass through the intersection within the prescribed cycle length and the given geometry of the links and intersection. The demand discharge is compared with the actual discharge by the feedback control loop to obtain the number of queued vehicles for the following time period. The number of queued vehicles is adjusted to include the demand for the next time period to better serve the demand for subsequent cycles.

Thus, the feedback control system acts as a means of evaluating the ability of SCATS to discharge the vehicle demand and ensures reasonable accuracy to demand estimates for subsequent time periods. Thus, the proposed integrated feedback control within the SCATS system enhances the performance and accuracy of the existing adaptive control system while still functioning within the operational framework of SCATS.
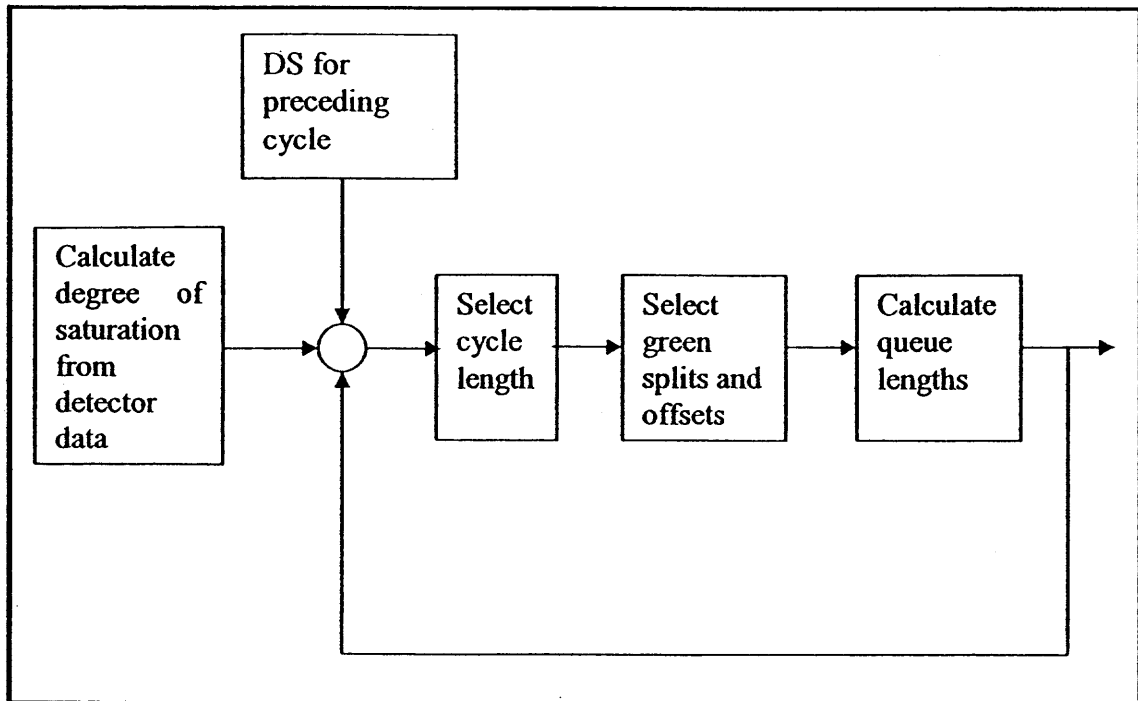
Figure 3.9 Details of feedback control system within SCATS.

3.11 Problem Statement in OPAC

The OPAC adaptive traffic control system uses a rolling horizon technique to reduce the shortcomings of prediction algorithms. The prediction algorithm used in previous models of OPAC predicted vehicle arrivals up to 1-2 minutes in advance of the actual arrival. This leads to inaccuracy in predictions because volume conditions might change during this time. The rolling horizon technique is used to reduce the prediction time of 1-2 minutes. The first head data measurement for this technique is obtained from detectors and is accurate. However, the process does induce some inaccuracy due to the lack of an evaluation mechanism.

In the rolling horizon process, the first roll period is operated based on the optimized policy calculated through the optimization algorithm, Optimal Sequential

Constrained Search (OSCS). Based on the calculations of this optimized plan, the vehicle arrival at the end of this roll period is obtained, and a new optimization strategy is calculated for the next roll period. Thus, an algorithm calculates the arrivals for the head of the next rolling period based on the optimized timing plan for the previous rolling period. This process is further extended until the end of the stage.

There is no mechanism, however, to compare the actual arrivals with the arrivals calculated using the prediction model. The timing plan for the entire stage is based on predicted arrival flows, and detector data is only used at the beginning of the stage. For the process to be implemented online, the calculation for the entire stage or horizon has to be made at the beginning of the process at the head of the first rolling period. If the actual vehicle flow does not accurately follow the expected pattern during any rolling period due to external factors, the optimization procedure for the next rolling period is not truly optimal, and the error is carried forward in time to cause further inaccuracies during the entire stage.

Further, each stage operates using the arrivals obtained at the start of a horizon, and the vehicles queued due to inaccuracies explained above are not considered for the next stage. Therefore, the detector data obtained at the head is not a correct representation of the demand in a link in case of queuing. There is a need for the evaluation mechanism to verify the values obtained at the tail of the roll period with the actual discharge values.

3.12 Integrating Feedback Control in OPAC

The above mentioned condition can be improved with the help of an estimation mechanism that can prevent the errors from propagating in time. Such an estimation mechanism can be devised with the help of a feedback control system, which can detect and correct errors introduced during the optimization process in OPAC.

Figure 3.10 represents the operation of OPAC in a feedback control environment. The blocks represent the functions in the procedure. The process begins with the measurement of arriving vehicles at the start of the first rolling period at the beginning of a stage n. According to the concept of rolling horizon technique, this data is used to calculate the optimized signal timing plan for the entire stage by the OSCS method. The optimized plan is then implemented only for the rolling period, and the vehicles that will reach downstream of the first rolling period will again be used to calculate the optimized plan for remaining stage length. Thus, the sequentially optimized plan is calculated for the entire stage. The sequential optimization procedure provides the output expected at the end of the stage. The proposed system incorporating feedback control into OPAC acts as an evaluation mechanism to compare the difference between the expected and actual outputs.

The current mechanism of OPAC considers the queue lengths only to calculate delays. The feedback control mechanism can make finer adjustments to the input volumes by adjusting the signal plan before the implementation stage so that the errors are not propagated to subsequent stages.
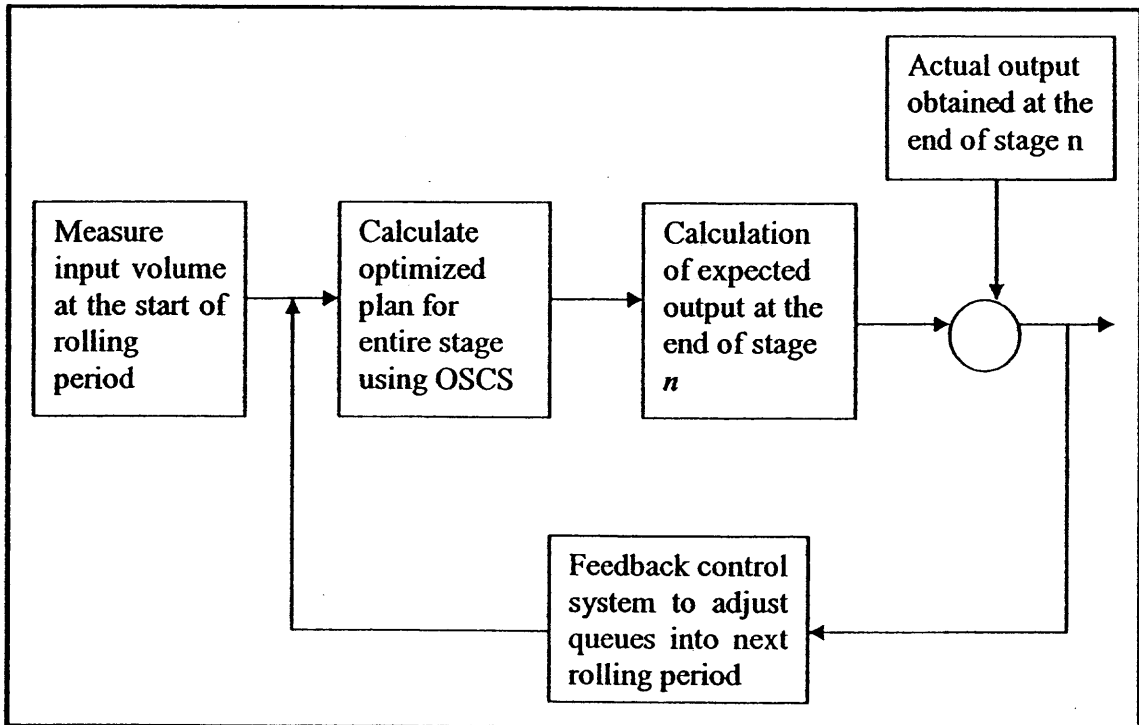
Figure 3.10 Conceptual framework representing OPAC in feedback environment.

Figure 3.11 is the block diagram representing the components of the feedback control system to enhance the performance of OPAC. The system is composed of a block which calculates the optimized signal timing plan for a stage n using OSCS. The block also calculates the corresponding discharge at the end of stage n using the rolling horizon technique. This calculated discharge is compared with the actual discharge obtained at the end of the stage. The difference between the expected and actual discharge gives the queue length for stage n. The queue length at stage n is adjusted during the beginning of the next stage and an optimized signal timing plan is calculated based on this demand.
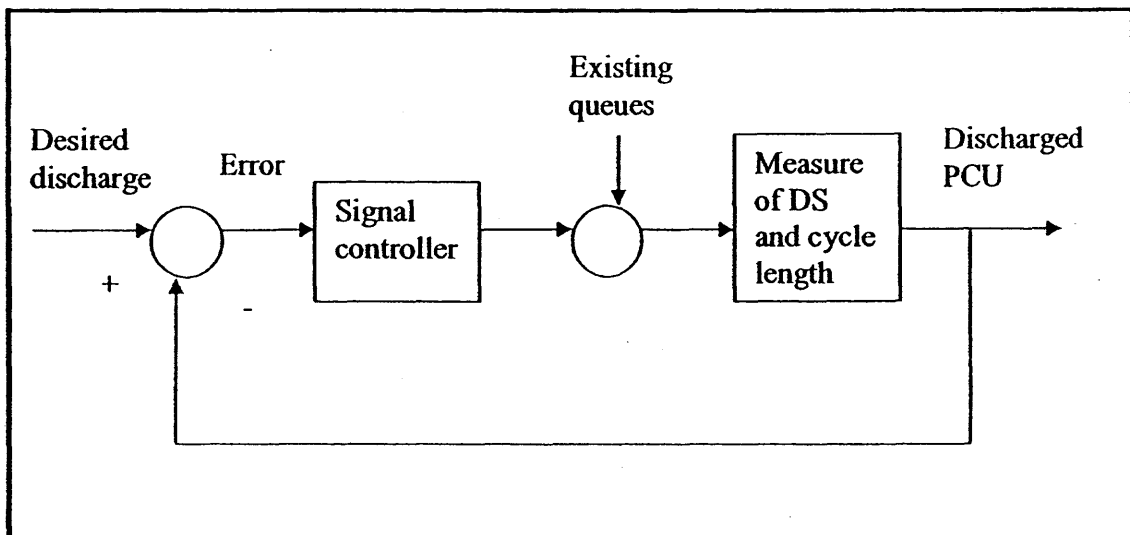
Figure 3.11 Details of feedback control system within OPAC.

3.13 Summary

This chapter dealt with the methodologies involved in the three leading adaptive traffic control strategies; SCOOT, SCATS and OPAC, and focused on the sources of errors in these systems which lead to their suboptimal performance. The chapter further discussed the proposed enhancement in the performance of these adaptive traffic control systems by operating them in a feedback environment and introduced the conceptual framework integrating feedback control with adaptive traffic control strategies. This was followed by details of basic feedback control systems, their elements and operational methodologies. Finally, a detailed description involving the application of the conceptual framework to the methodologies of SCOOT, SCATS and OPAC was provided and the expected enhancements were discussed.

CHAPTER 4

MODEL FORMULATION

This chapter describes the mathematical formulation of the proposed feedback control system. It is beyond the scope of this research to verify the efficiency of feedback control systems in improving the performance of SCOOT, SCATS and OPAC because this would either require development of simulation models implementing adaptive control strategies, or field implementation of control strategies. Hence, the performance of a hypothetical arterial of two intersections has been tested under the effect of feedback control and the performance has been subsequently evaluated through a detailed sensitivity analysis. This chapter involves the development of a traffic flow model for describing the flow of vehicles on the arterial. The development of the traffic flow model is followed by the mathematical design of feedback controllers and the formulation of expressions for controller parameters using transformation techniques. The formulation includes the design of Proportional and Proportional-Integral controllers. The formulation also describes the integration of feedback controllers with the traffic flow models for the intersections under consideration. Block diagrams for feedback control systems with two different types of controllers are also provided.

4.1 System Description

This section applies feedback control theory to develop an algorithm for control of traffic flow of two intersections within an arterial, while maintaining a desired output from the arterial. The research proposes a conceptual framework for integrating the concepts of

feedback control with existing adaptive traffic control systems to achieve better traffic management. This section attempts to support the proposed framework by providing a proof of the effectiveness of the concepts of feedback control in traffic management for a study arterial. Figure 4.1 represents the geometric layout of the study intersections and the direction of traffic flow. The traffic flow from an upstream to a downstream intersection is modeled under a feedback control setup.
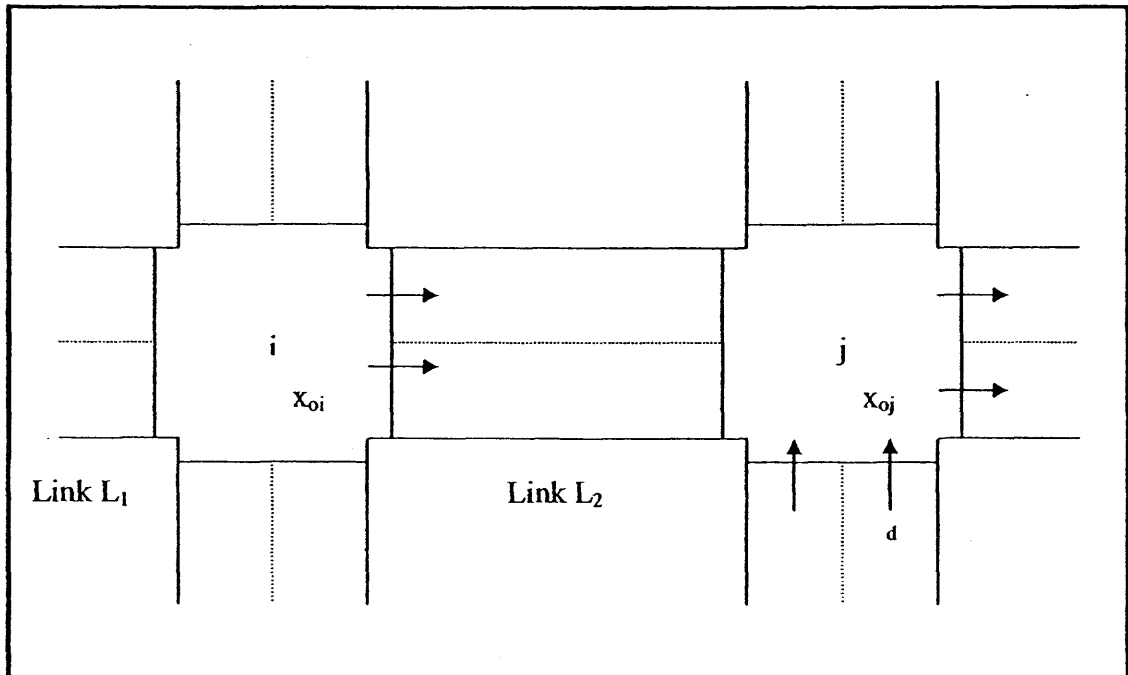
Figure 4.1 Geometric layout of the system under study.

4.2 Controller Design

The proposed control strategy is first designed for control at two adjacent intersections. The analysis can be extended to a larger arterial after the validity of the model is tested and satisfactory results are attained. This section describes the process for designing the controllers used in the feedback system being studied. The traffic flow models developed in this study use the principle of conservation, i.e. all the vehicles that enter the system at any time must leave the system at some point. The traffic flow models, which are

represented as difference equations, use transfer functions. The transfer functions establish a relation between the output of and input to the controllers. Depending on these transfer functions, a controller regulates the control variables. Two types of controllers are applied to the system, Proportional and Proportional-Integral controller.

4.2.1 Assumptions

The proposed control system is a linear time-invariant system. To apply this feedback system to intersections, the geometric details of the arterial under consideration are assumed to be known. The maximum queue holding capacity of the links and the saturation flow rates at the intersections are based on the geometric constraints. The maximum allowable green times are assumed to be known. The time lag for vehicles to travel from one intersection to another is not accounted for due to the assumption that links in the study network have short link lengths. A one-way arterial roadway is modeled with left turning movements on the cross-street not permitted; hence the only disturbance to the system is in the form of right turning vehicles and through vehicles arriving from the upstream intersection. The arterial is assumed to have two-lanes. No mid-block sources are considered. A fixed cycle length of 100 seconds is assumed but would vary depending upon the historical data for the intersection. A saturation flow rate of 1800 vphpl is used for all intersection conditions.

4.2.2 Notations

The following notation was used for the mathematical formulation in this chapter:

$x_{oi}(k)$ = Volume to be released from intersection "i" into link L_2 as shown in Figure 4.1, by the controller over a period $[k\Delta T, (k+1)\Delta T]$, veh/sec,

k = Time index that varies as 1, 2, 3...

$x_j(k)$ = Total queue holding capacity of link L$_2$,

$x_{oj}(k)$ = Volume released from link L$_2$ during a phase,

$x_{error}(k)$ = Error term giving the difference between the capacity at an intersection and the actual output,

c_j = Discharge capacity of intersection j,

$d_j(k)$ = Right turning movements on an intersection,

T = Sample time,

k = Sample in question,

K_p = Regulator parameter of the Proportional controller,

K_I = Regulator parameter of the Proportional-Integral controller.

The following section describes the development of controllers for the arterial network described above.

4.2.3 Design of Controllers

4.2.3.1 Proportional Controller. The Proportional controller used in this research controls the number of vehicles that need to be discharged from the upstream intersection to achieve maximum output from the downstream intersection. Thus x_{oi} the output volume, is the control variable. The P-regulator uses the following expression to determine the number of vehicles to be released from an upstream intersection to maintain a specific traffic volume condition at the subsequent downstream intersection:

$$x_{oi}(k) = x_{oi}(k-1) + K_P \left[c_j - x_{oj}(k) \right]$$

(4.1)

$$x_{error}(k) = c_j - x_{oj}(k) \qquad (4.2)$$

$$x_{oi}(k) = x_{oi}(k-1) + K_P[x_{error}(k)] \qquad (4.3)$$

Equation (4.1) indicates that the number of vehicles discharged from intersection "i", the upstream intersection is dependent on the number of vehicles to be discharged by the next downstream intersection. The proportional regulator controls the discharge of vehicles from intersection "i" into link L_2. The difference between the capacity of intersection "j" and the actual output from the intersection represents the error of the system. Equation (4.2) describes the error of the system represented as X_{error}. Substituting Equation (4.2) in Equation (4.1), we obtain Equation (4.3) which represents the control law for the study problem.

Equation (4.3) uses a regulator parameter K_p to adjust the inflow of vehicles from upstream. The regulator parameter acts as a coefficient or correction factor which is applied to the "error" in order to obtain the upstream input volumes.

In the presence of turning movements at a downstream intersection, the total number of vehicles discharged from the upstream intersection will be affected. This is true because the total number of vehicles to be released into the downstream intersection is the summation of incoming vehicles from upstream and the turning vehicles. Thus, the outflow of vehicles from intersection "j" is given as follows:

$$x_{oj}(k) = x_{oi}(k) + d_j(k) \qquad (4.4)$$

where, $d_j(k)$ represents the right turning movements into intersection j for time period k. For the previous time period, the same equation is written as

$$x_{oj}(k-1) = x_{oi}(k-1) + d_j(k-1) \qquad (4.5)$$

The difference between the outputs for two subsequent time periods has to be reduced by the controllers in order to reduce the error of the system.

Depending on the value of this difference, the controller calculates an appropriate value of the controller parameter. The difference in the output values is used by the controller to decide whether to increase or decrease the input from upstream with respect to the current input value. The difference in the output of two subsequent time periods is obtained by subtracting the expressions representing the outputs for two time periods. Thus, subtracting Equation (4.5) from Equation (4.4), the following expression is obtained.

$$x_{oj}(k) - x_{oj}(k-1) = x_{oi}(k) - x_{oi}(k-1) + d_j(k) - d_j(k-1) \tag{4.6}$$

Substituting equation (4.2) and (4.3) in the above expression, Equation (4.6) can be represented in terms of control parameter as follows:

$$x_{oj}(k) - x_{oj}(k-1) = K_p[c_j - x_{oj}(k)] + d_j(k) - d_j(k-1) \tag{4.7}$$

To obtain the transfer function for the "plant" of the control system, which represents the traffic flow model, z-transfer function of the control law is determined. The z-transfer function for the control law given by Equation (4.3) can be calculated as

$$X_{oi}(z) = z^{-1} X_{oi}(z) + K_P X_{error}(z) \tag{4.8}$$

$$X_{oi}(z)(1 - z^{-1}) = K_p X_{error}(z) \tag{4.9}$$

$$\frac{X_{oi}}{X_{error}} = \frac{zK_p}{z-1} \tag{4.10}$$

$$H(z) = \frac{X_{oi}(z)}{x_{error}(z)} = \frac{zK_p}{z-1} \tag{4.11}$$

Equation (4.7) represents the flow equation for the difference between outputs from intersection j for time period $k-1$ and k. The equation is derived for a control system with a proportional regulator. To obtain the effect of disturbance d_j and the output from intersection j, a general flow equation can be obtained from Equation (4.7). Without the inclusion of a regulator, i.e. $K_p=1$, the general flow equation is given as

$$x_{o_j}(k) = x_{o_j}(k-1) + \left[c - x_{oj}(k)\right] + d_j(k) - d_j(k-1) \tag{4.12}$$

The z-transfer function for the feedback law given by Equation (4.12) is given as

$$X_{oj}(z) - z^{-1}X_{oj}(z) - X_{oj}(z) = d_j(z) - z^{-1}d_j(z) \tag{4.13}$$

$$X_{oj}(z)(1 - z^{-1} + 1) = d_j(z)(1 - z^{-1}) \tag{4.14}$$

$$\frac{X_{oj}(z)}{d_j(z)} = \frac{1 - z^{-1}}{1 - z^{-1} + 1} \tag{4.15}$$

$$\frac{X_{oj}(z)}{d_j(z)} = \frac{z - 1}{2z - 1} \tag{4.16}$$

Equation (4.12) indicates the effect of the disturbance d_j on the output x_{oj}. The effect is represented in terms of a z-transformation by Equation (4.16). Similarly, the effect of upstream incoming volume x_{oi} on the output can be obtained by the z-transform of Equation (4.6) considering a constant disturbance d_j. The combined effect of both upstream volumes and the disturbance is given by combining the two effects.

$$\frac{X_{oj}(z)}{X_{oi}(z)} = 1 \tag{4.17}$$

Thus, the combined effect of the two incoming volumes on the output is given as

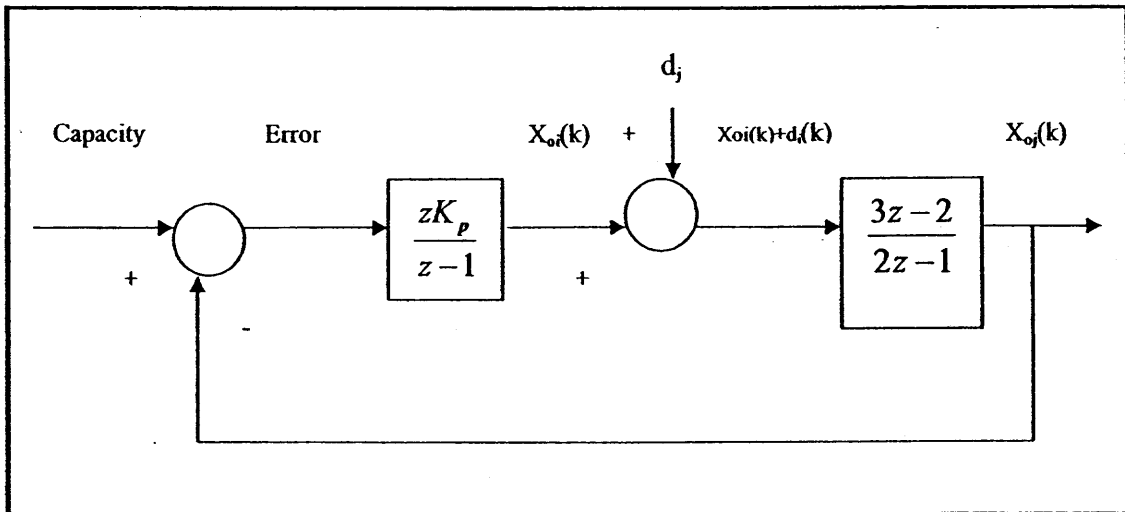$$H(z) = \frac{3z - 2}{2z - 1} \tag{4.18}$$

Figure 4.2 Block diagram for P controller.

The block diagram representing the feedback control structure using the proportional controller is shown in Figure 4.2. The figure shows a representation of all the components of the proposed feedback control system. The figure indicates that the output from the controller depends on the expression given by Equation (4.11). The controller, depending on the error, decides the input volume that should be allowed into the downstream intersection. Thus, the transfer function of the controller, which is represented as the z-transform of traffic flow differential equations, controls the demand for the following time period i.e., it updates the demand for each cycle. The signal timing plan is calculated based on the updated demand.

4.2.3.2 Proportional-Integral Controller. The transient behavior of the closed loop system can be improved by using a PI- regulator. Applying a PI-regulator to the control loop gives the following feedback law

$$x_{oi}(k) = x_{oi}(k-1) + K_p[x_{error}(k) - x_{error}(k-1)] + K_I[x_{error}(k)] \qquad (4.19)$$

The z-transfer functions for the above feedback law of regulator is calculated as

$$X_{oi}(z) - z^{-1}X_{oi}(z) = K_p\left[X_{error}(z) - z^{-1}X_{error}(z)\right] + K_I[X_{error}(z)] \qquad (4.20)$$

$$H(z) = \frac{X_{oi}(z)}{X_{error}(z)} = \frac{z(K_I + K_p) - K_p}{z - 1} \qquad (4.21)$$

Since the change in the regulator does not affect the basic equation of traffic flow, the transfer function for of the equation describing the traffic flow within the feedback loop remains the same, and only the transfer function for the regulator changes to Equation (4.21).

Figure 4.3 illustrates the control system with a proportional-integral controller. The figures describe the traffic flow process with the transfer function of the controllers described in the algorithm.
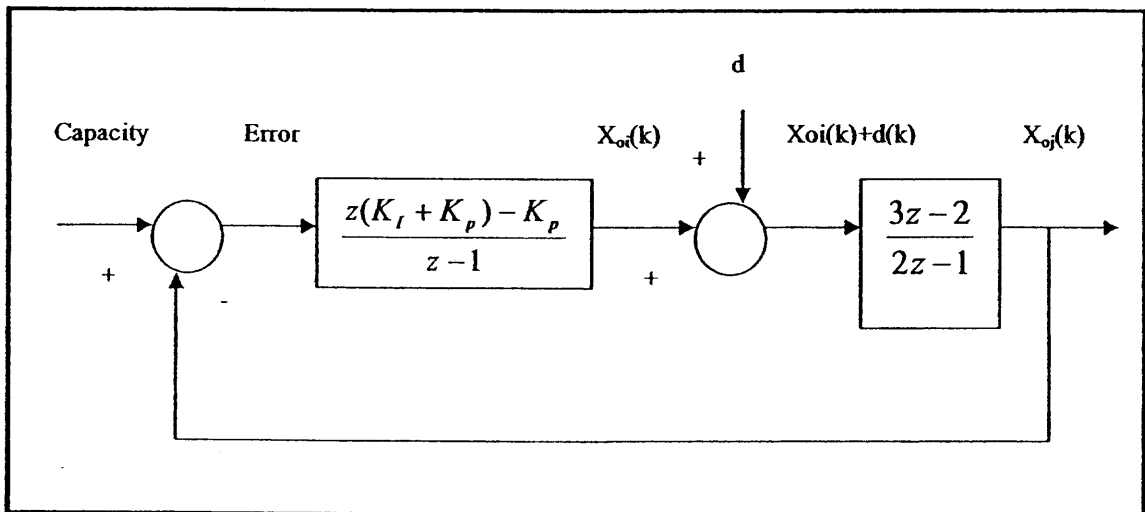

Figure 4.3 Block diagram for PI controller.

The derivations shown above do not consider travel time between intersections. The feedback control system designed in this research attempts to provide a corrective mechanism for the estimation procedure of adaptive traffic control systems. The control system is not designed to consider the delays associated with left turning vehicles or parking maneuvers within links. Moreover, these factors are already considered in the

traffic flow models of the adaptive control systems like SCOOT. Thus, it can be reasonably assumed that the travel time of vehicles between links is negligible.

The output volume from the downstream intersection can be used to obtain sufficient green time to be used by the strategy controlling the intersections. The following relationship will be used to convert the output volumes to corresponding green times:

$$g_i(k) = \frac{x_{oi}(k)}{S_i} \frac{C}{D_i}$$
(4.22)

$$g_i(k) < g_{i,max}$$
(4.23)

$x_{oi}(k)$ = Demand volumes for link i to be served for time period k;

$g_i(k)$ = green time provide to the approach;

C = cycle length;

D_i = degree of saturation; and

S_i = saturation flow rate.

The degree of saturation is used in the no-control case for obtaining over and undersaturated volume conditions. The no-control case refers to the base case that will be described in Chapter 5. The base case is used to evaluate the performance of a feedback loop without the inclusion of controllers. In the absence of controllers, the only regulating device to control the inflow volumes from the upstream is the capacity and maximum and minimum allowable green times. The base case has been studied to evaluate the efficiency of controllers in the feedback control systems to achieve better queue management by comparing them with a simple feedback loop with no controllers.

4.2.3.3 Selection of Controller Parameters (K_p and K_i)

Controllers influence the feedback controller systems depending on their individual characteristics (Refer to Appendix A). The output from a feedback controller depends on the value of the controller parameter selected for a system. In theory and in practice, most feedback controllers do not achieve a zero error at the steady state. Hence, to obtain the most appropriate regulator parameter value, a tolerance level for the error should be set. Usually, a tolerance level within 2 to 5% of the steady state is used. Considering a steady state error of zero for the study system, a regulator parameter value which provides an error value within $\pm$ 0.05 from the steady state error is considered allowable for a P control system. Similarly, for a PI control system, the allowable tolerance level of ±0.03 is considered for selecting the best regulator parameter values.

4.3 Extension to Larger Arterials

This study can be extended to serve multiple intersection arterials or networks. The system requirements for larger arterials or networks would require that each intersection be considered as a separate control system with a feedback controller, i.e. each intersection be controlled by a separate feedback controller. The assumption of one-way streets would also need to be relaxed for large arterials and networks.

Intersections within an arterial can have varying physical characteristics. Depending on these characteristics, each intersection would possess varying throughput and queue capacities. Under the constraint of different throughput capacities, the individual feedback controllers would control the traffic volume to obtain the individual desired throughput. In this case, coordination between demands of each intersection would be required to obtain an optimal performance for the entire arterial or network. A

signal coordination plan would be required for larger intersections. However, each individual intersection would show optimum results subject to the constraint of available input from its upstream intersection.

CHAPTER 5

RESULTS AND ANALYSIS

5.1 Introduction

From the discussion of adaptive control system procedures in Chapters 1, 2 and 3, it has been established that the basic framework of some of the existing adaptive control systems considered in this research function on a feed-forward concept of data flow. Since flow of traffic is a dynamic phenomenon, adaptive control systems are expected to respond to field traffic conditions in real time. Feedback control has been proposed as a component, which when added to the basic framework of an adaptive control system, enables existing adaptive control systems to better update certain initial parameters to ensure the desired performance of the system.

The following section illustrates the application of feedback control in an adaptive control setting. A simulation tool, MATLAB, is used to find solutions to demonstrate the control system environment. The controller parameters along with the procedure followed by the feedback loop to find the solution are also provided. The research also considers the effect of different volume conditions on the performance of the control system using two different controllers. The effectiveness of the controllers to achieve better queue management with and without a controlled feedback loop is also investigated.

5.2 Simulation

It is out of the scope of this research to test the improvement in the performance of SCOOT, SCATS and OPAC due to the incorporation of a feedback control loop as this would require either a field implementation of these control strategies or the development of a simulation model capable of implementing adaptive control strategies. As a result, it was necessary to test the performance of a roadway network in a feedback environment with the help of MATLAB simulation. This section describes the simulation and measures of effectiveness considered in this study. An arterial of two intersections is further simulated under optimal signal timing plan within TRANSYT and MATLAB. The results of the two simulation techniques are compared to prove the efficiency of MATLAB in better handling queues in congested conditions.

5.2.1 Simulation Techniques

The control system in this study was simulated using the computer program MATrix LABoratory or MATLAB (Version 6.0.0.88 Release 12). SIMULINK is the tool within MATLAB that performs the simulation with the help of inbuilt interactive block diagrams representing functions. MATLAB has been used for simulation in this study since it allows a large number of inputs to be fed into the control system in order to evaluate the performance of the proposed feedback control under a wide range of conditions. The program is capable of simulating complex feedback or non feedback processes and has the flexibility to handle transformations and a variety of input patterns to suit real life industrial processes. Also, MATLAB has the capacity to handle state space equations and z-transforms, which constitute an important part of the design of

controllers. The direct use of PI or PID controllers when the values of the regulator parameters are already known is also possible in MATLAB.

The following sections of this chapter demonstrate the simulation of a two intersection arterial using MATLAB. The simulation illustrates the steps followed by the feedback control system to regulate the control parameter of the system, i.e. the input vehicles from the upstream to achieve a desired outflow from the downstream. Based on the mathematical formulations of the control strategy described in Chapter 4, the simulated control system obtains the most suitable values of regulator parameters for the most efficient queue management within the arterial under study.
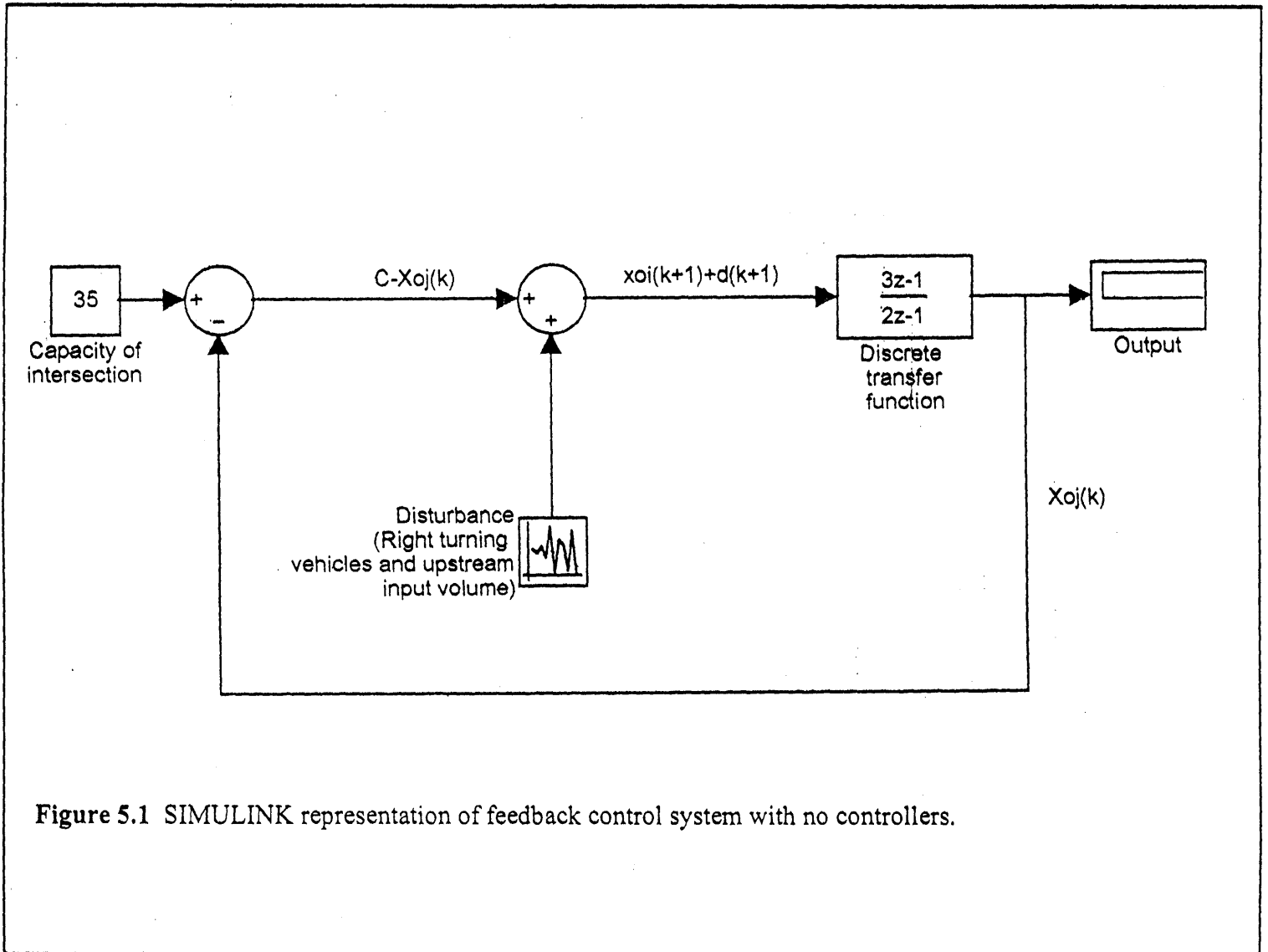
5.2.2 Measures of Effectiveness

The measures of effectiveness (MOE) used to evaluate the performance of the control strategy are the steady state error and throughput from the system. The steady state error describes the difference between the desired and actual outputs from the system. The throughput from the system describes the number of vehicles discharged from an intersection within each cycle. Since the output determines the ability of the system to stabilize the variable input, the output from this system when compared to the input, is considered as a measure of effectiveness. Moreover, the steady state error is a measure of effectiveness of the control strategy as this measure demonstrates the ability of the strategy to bring the control system to steady state in a minimum amount of time. This MOE demonstrates the efficiency of the system to track the desired throughput from an intersection.

5.3 Solution Process

The following describe the steps to implement the control strategy and stabilize the feedback control system under variable inputs. For the network of two intersections described in section 4.1, with specific geometric and volume conditions, the algorithm consists of the following steps:

1) Inputs in the control system, including the throughput and error or difference between the desired and actual throughput, are initialized to zero.

2) The predetermined value of the desired capacity from the system is fed into the system as a constant.

3) Input into the system is in the form of a random uniform number ranging from 0 to 65 vehicles per green (vpg), which represents the number of vehicles to be served at the downstream intersection. Volumes are input into the system in the form of turning vehicles and vehicles from the upstream intersection (refer to Figure 4.1). The equilibrium of the system is disturbed when there is a difference between the desired output volume, which is the capacity of the downstream intersection, and actual output. This disturbance is used in the feedback loop to stabilize the system by reducing the steady state error.

4) The input is fed through each module of the feedback loop. First it passes through the module that determines the output based on the capacity constraint of 35 vpg. Output from this module represents the throughput from the downstream intersection (refer to Figure 5.1). This throughput follows the feedback loop, and is compared to the constant predetermined capacity.

Figure 5.1 SIMULINK representation of feedback control system with no controllers.

In the figure, the blocks and signals are labeled as follows:

- A block labeled **35** with caption **Capacity of intersection**
- A summing junction with **+** and **−** inputs, output signal **C-Xoj(k)**
- A summing junction with **+** and **+** inputs, output signal **xoi(k+1)+d(k+1)**
- A disturbance block captioned **Disturbance (Right turning vehicles and upstream input volume)**
- A discrete transfer function block $\dfrac{3z-1}{2z-1}$ captioned **Discrete transfer function**
- An **Output** block
- Feedback signal labeled **Xoj(k)**

5)   A comparison of the capacity and the output determines the error in the system. If the capacity is higher than the input, there is no error, since all the volume that enters into the system will exit without any queue buildup. In case the turning volumes are higher and result in an input equal to the capacity, the number of through vehicles that can be allowed from the upstream intersection will be reduced. This will lead to a negative error, which indicates a queue buildup on the main street.

6)   In the absence of a queue, the error value will be positive, and the input volume from the upstream can at least be equal to the capacity of the intersection for the next time period. But, if the error value is negative which indicates the presence of a queue, the input from the upstream intersection has to be less than the capacity. Thus the error determines the input volume from the upstream intersection for the next time period, hence increasing or decreasing the total input as compared to the previous time period. This is the new input to be fed into the system for the next time period.

7)   The time taken to stabilize the system, i.e. bringing the system to a steady state, is plotted against the MOEs of the system, such as steady state error or throughput from the system. The time taken by the system to reach a steady state is an indicator of the efficiency of the system. Hence, a shorter time taken to stabilize the system under given values of regulator parameters indicates a better system. The input and output can also be compared to see how efficiently the system stabilizes random values of inputs.

The proposed control system in its present form is sufficient to handle both over and undersaturated volume conditions at intersections. The system is based on metering of vehicles at the upstream intersection. During oversaturated conditions, the output from the system is equal to the constant capacity since the system cannot serve more than

capacity. For undersaturated conditions, when the sum of the incoming volume to be served and the error or queue from the previous time index is lower than the capacity, then the throughput from the system is the sum of these two quantities i.e. the total volume to be served. This function is carried out in MATLAB with the help of a switch which compares the current demand with the capacity of the system and decides what should be the throughput of the system.

5.4 Application of Feedback Control Approach

The mathematical model described in Chapter 4 describes the flow of vehicles within the proposed feedback loop. To check the effectiveness of the controllers in the system, simulation was carried out for two systems in MATLAB. The first simulation was carried out without the application of a controller and the second simulation was carried out with the inclusion of the controller. The comparison demonstrates the ability of well designed feedback controllers in better handling demand as compared to a feedback system without controllers. The arterial is provided all the remaining green after servicing the cross street. The capacity of the intersection for a cycle length of 100 seconds is equal to 35 vehicles per green. A variable input volume range of 0 to 65 vehicles per green is provided to the intersection during the simulation period to accommodate both under and oversaturated conditions.

5.4.1 No-Control Approach

In the no-control approach, there is flow of data within a feedback loop without a controller in the loop, i.e. all the vehicles entering the system leave the system without being acted upon by any external factors and follow the constraints of maximum green.

Thus, the controller acts like a fixed-time signal. The property to adapt to changing inputs and track the desired output is achieved by the introduction of a feedback loop. The measures of effectiveness are the same as described above. The SIMULINK representation of the system is shown in Figure 5.1. The figure represents a block diagram of a feedback control system applied to an arterial without including controllers. Figure 5.2 shows a comparison of the total input volume to the output of the system. It is evident from Figure 5.2 that the output from the feedback system under the no-control approach tries to track the variations in the input volumes. This indicates the adaptability of the feedback system to changing inputs. Figure 5.3 shows the variation in the error values of the system for different input volumes. The figure shows a comparison of the desired error value, i.e. zero, and the actual error induced in the loop with the changing inputs in terms of turning volumes. The figure indicates a variation in error ranging from 0 to 17 vpg, which is a high value of error. The error can be reduced with the help of different types of controllers as discussed in the next section. Figure 5.4 represents the actual output of the system compared to the desired output. The figure indicates that in the absence of controllers, the actual output obtained from the intersections is less than the capacity during the simulation period. This explains a need to include a well designed controller into the feedback control system.
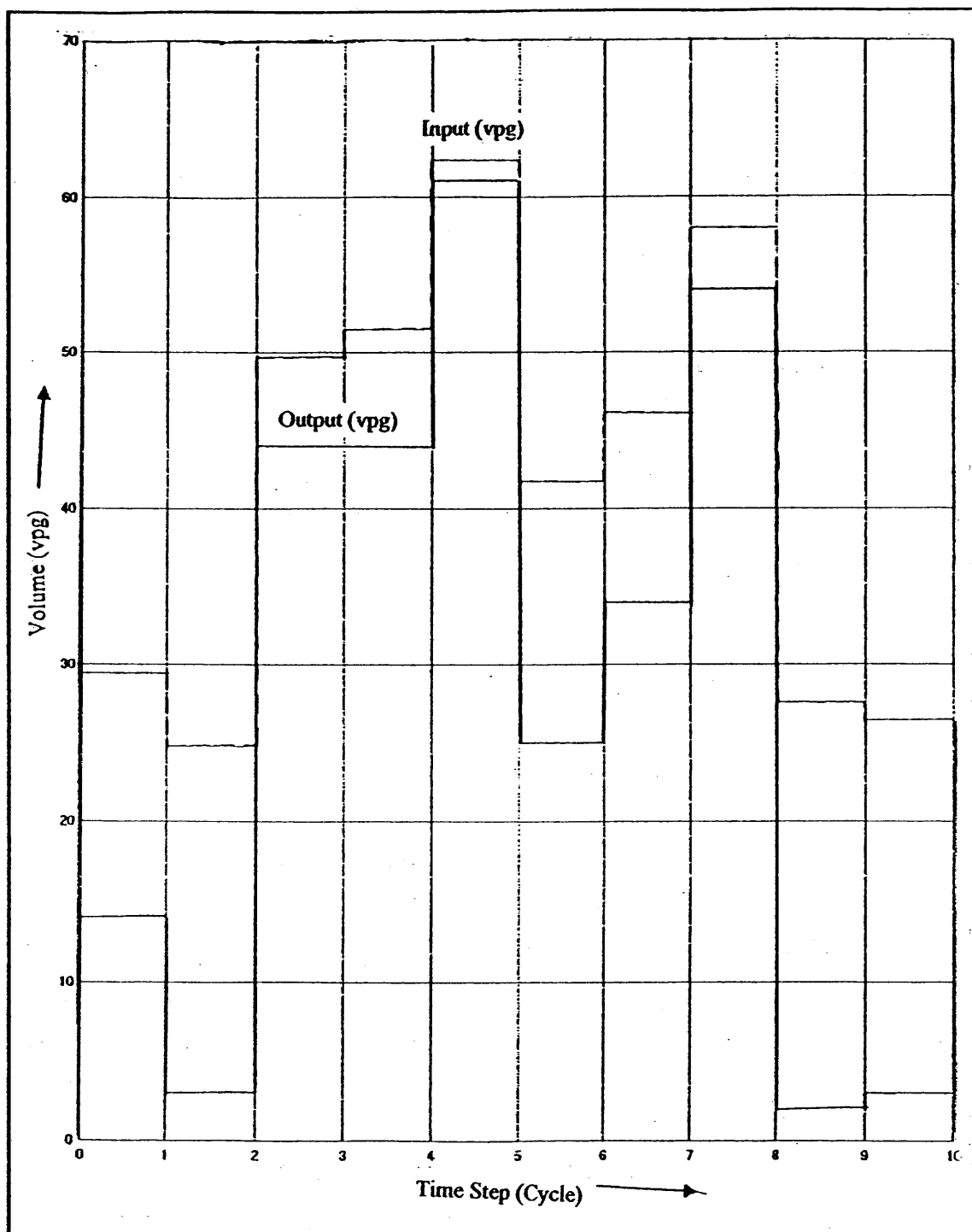
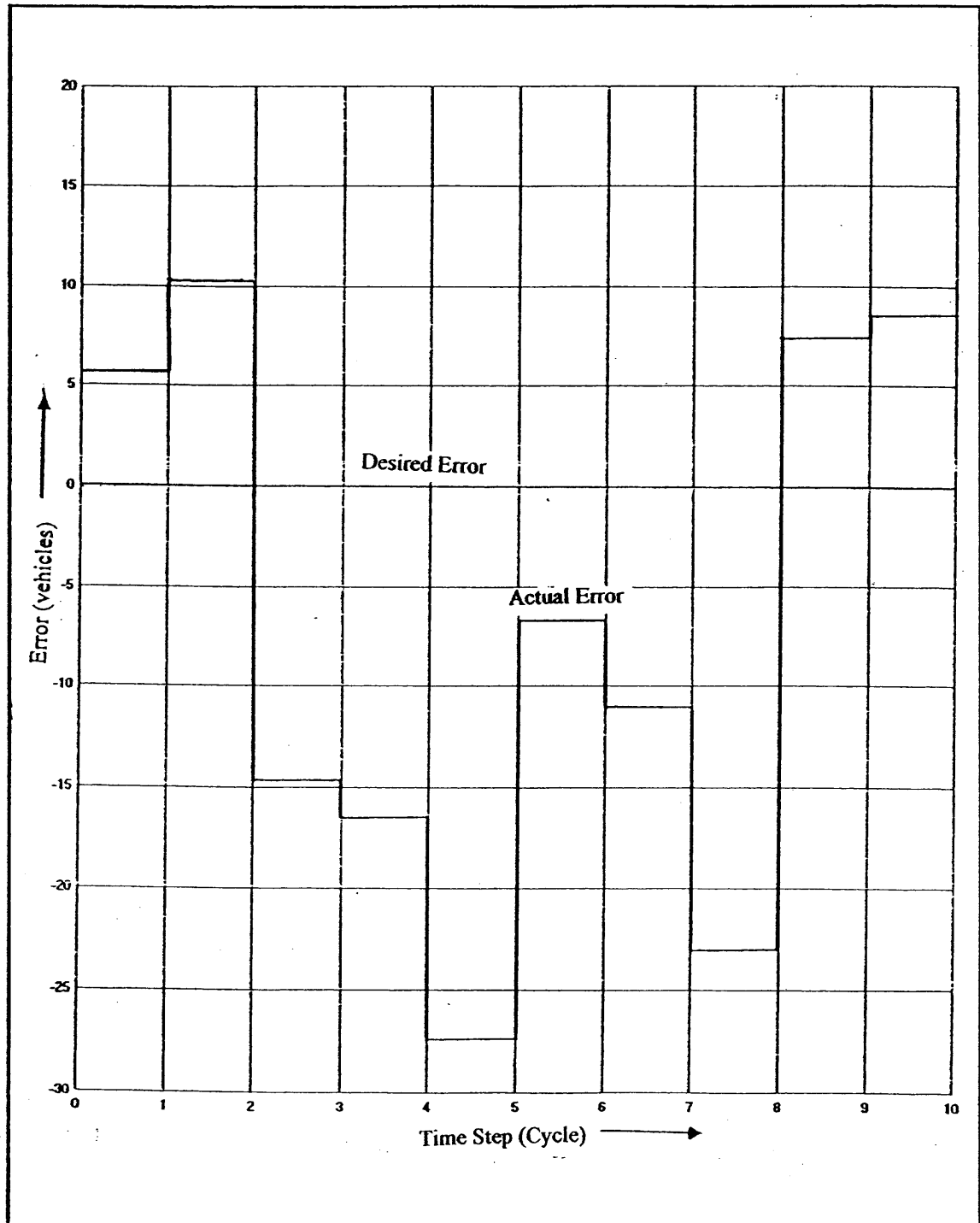**Figure 5.2** Comparison of total input and output for no-control case.

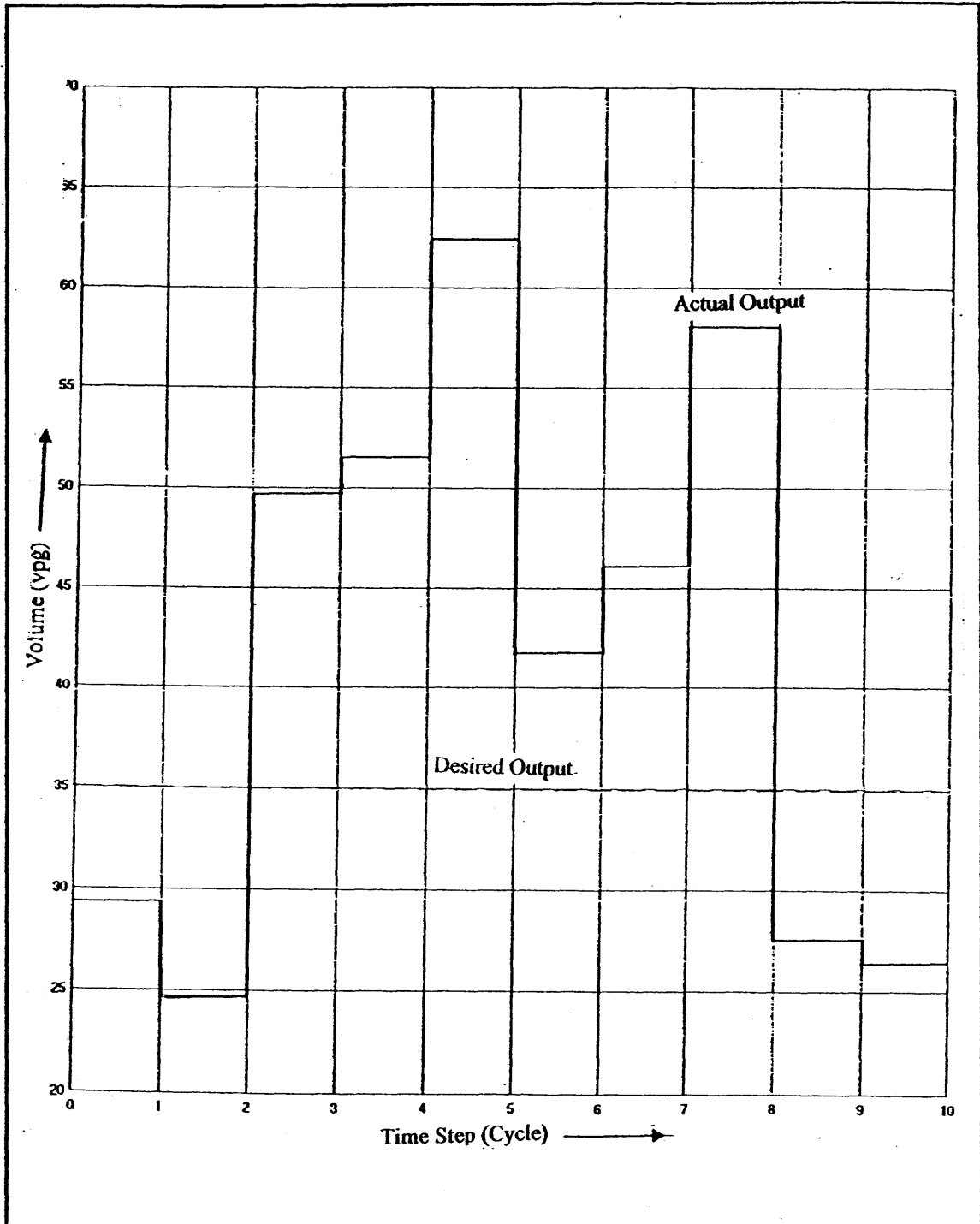**Figure 5.3** Comparison of actual and desired error for no-control case.

Figure 5.4 Comparison of actual and desired output for no-control case.

5.4.2 Controlled System Approach

The controlled system approach uses feedback loop to regulate the controller. In this research, the controllers applied are Proportional (P) and Proportional-Integral (PI). The effect of both controllers on the feedback system is studied. To carry out the sensitivity analysis of the system under the effect of two types of controllers, the system is tested with changing input volumes. The results from this sensitivity analysis are compared to the results of the no-control case.

Case 1. Proportional Control – Constant Volume

Figure 5.5 is a SIMULINK representation of a feedback control system with a proportional regulator for a given capacity of 35 vpg for the main street and a constant total input volume of 65 vpg. The block diagram represents a set desired capacity of 35 vpg, a constant disturbance, a P-regulator, discrete transfer function relating the output with uninterrupted through vehicles from upstream and right turning vehicles, output from the downstream intersection and finally, the block comparing desired and actual error. Results of the control system comparing the actual and desired error are provided in Figure 5.6. In the first time step, the error is 0.06. The figure indicates that the system reduces its errors gradually and finally attains an error value close to the desired error of 0 after 16 time units of simulation, which is equal to 16 cycle lengths. During the entire simulation period of 16 cycle lengths, a constant input volume of 65 vpg is fed into the arterial, which constitutes a highly oversaturated condition. The feedback control system constantly updates the demand and meters the input volume from the upstream accordingly to bring the queue lengths to a minimum.
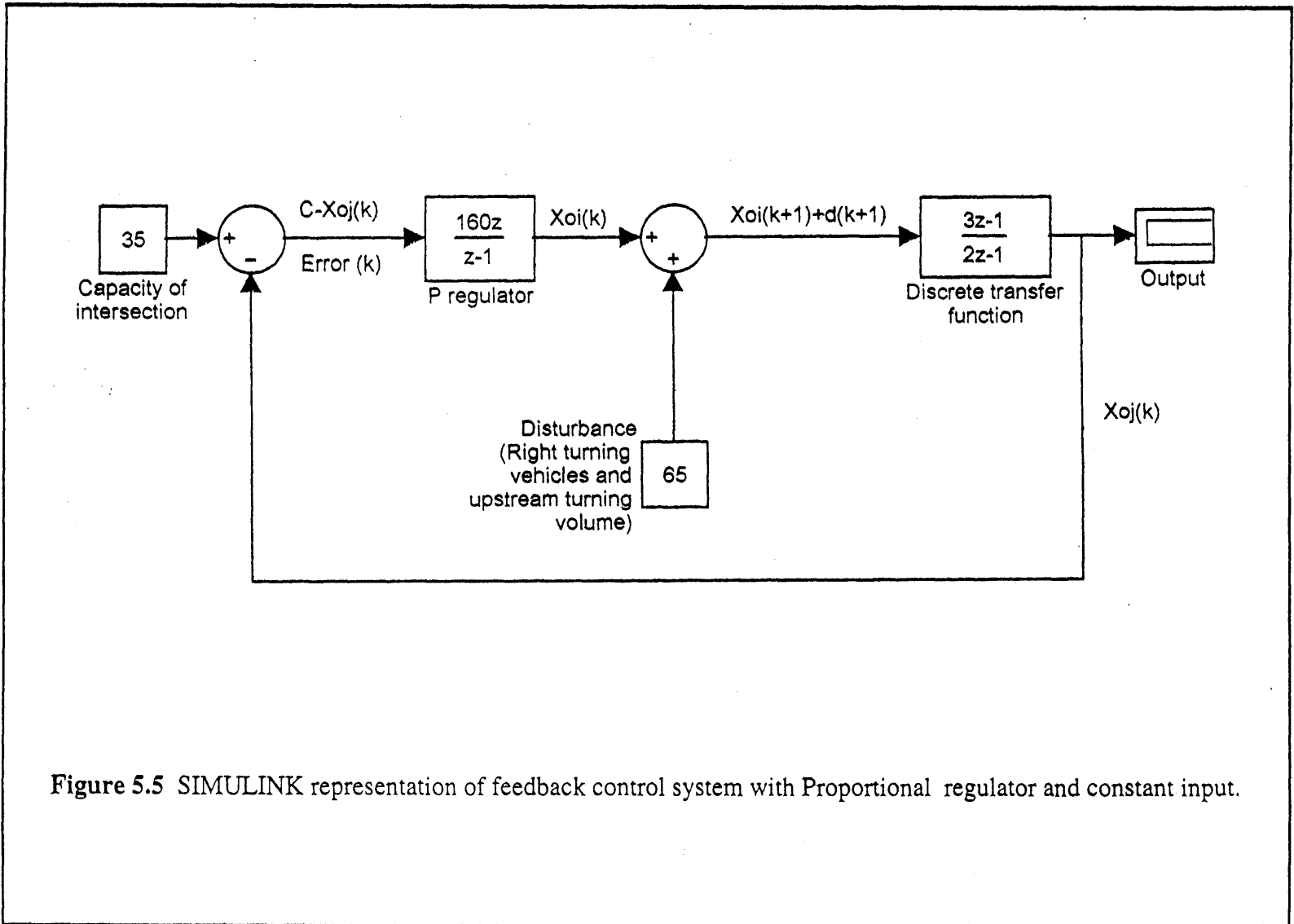
Figure 5.5 SIMULINK representation of feedback control system with Proportional regulator and constant input.
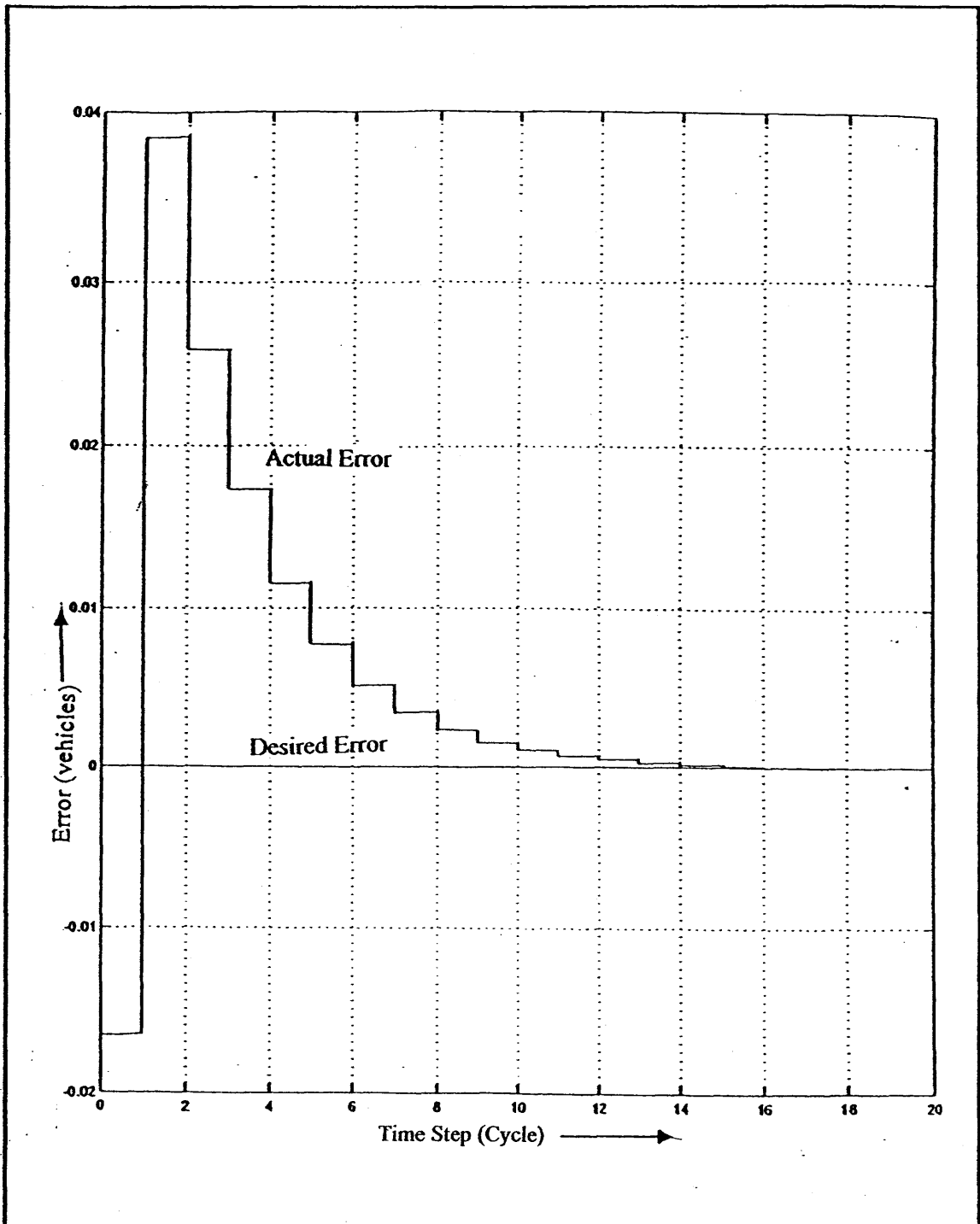
Figure 5.6 Comparison of actual and desired error for P-control system with constant input.

Similarly, a comparison of actual and desired output for the same system is represented in Figure 5.7. The figure shows that initial periods of simulation have lower output compared to the capacity. The desired output of 35 vehicles per cycle is obtained after 16 time units, i.e. 16 cycle lengths when the error of the system approaches zero. Thus the capability of the feedback control system to bring the errors close to zero is proved.

Case 2. Proportional Control-Variable Volume

Figure 5.8 shows the SIMULINK representation of a system with a proportional regulator for a given capacity of 35 vehicles per cycle for the main street and a total input volume varying between 0 and 65 vehicles. These volumes include the turning volume from the cross street and the upstream incoming volume.

Figure 5.9 shows the comparison of the inflow and outflow from the control system. The figure indicates that the controller stabilizes the random input into the control system and smoothes out the fluctuations of the input, thus causing a relatively smoother outflow of vehicles from the intersection. Figure 5.10 compares the desired and actual error of the system. The figure indicates that the error in the system is large in the simulation, but generally decreases over time. Figure 5.11 is a comparison of the desired and actual output from the system. Compared to the analysis of the case where constant input volumes are used in the simulation, the errors do not converge to the desired error. This is due to the fact that the control system tries to track the input for one simulation period. Since the input volume changes every time period, the system is unable to reach a stable state that is represented by zero error.
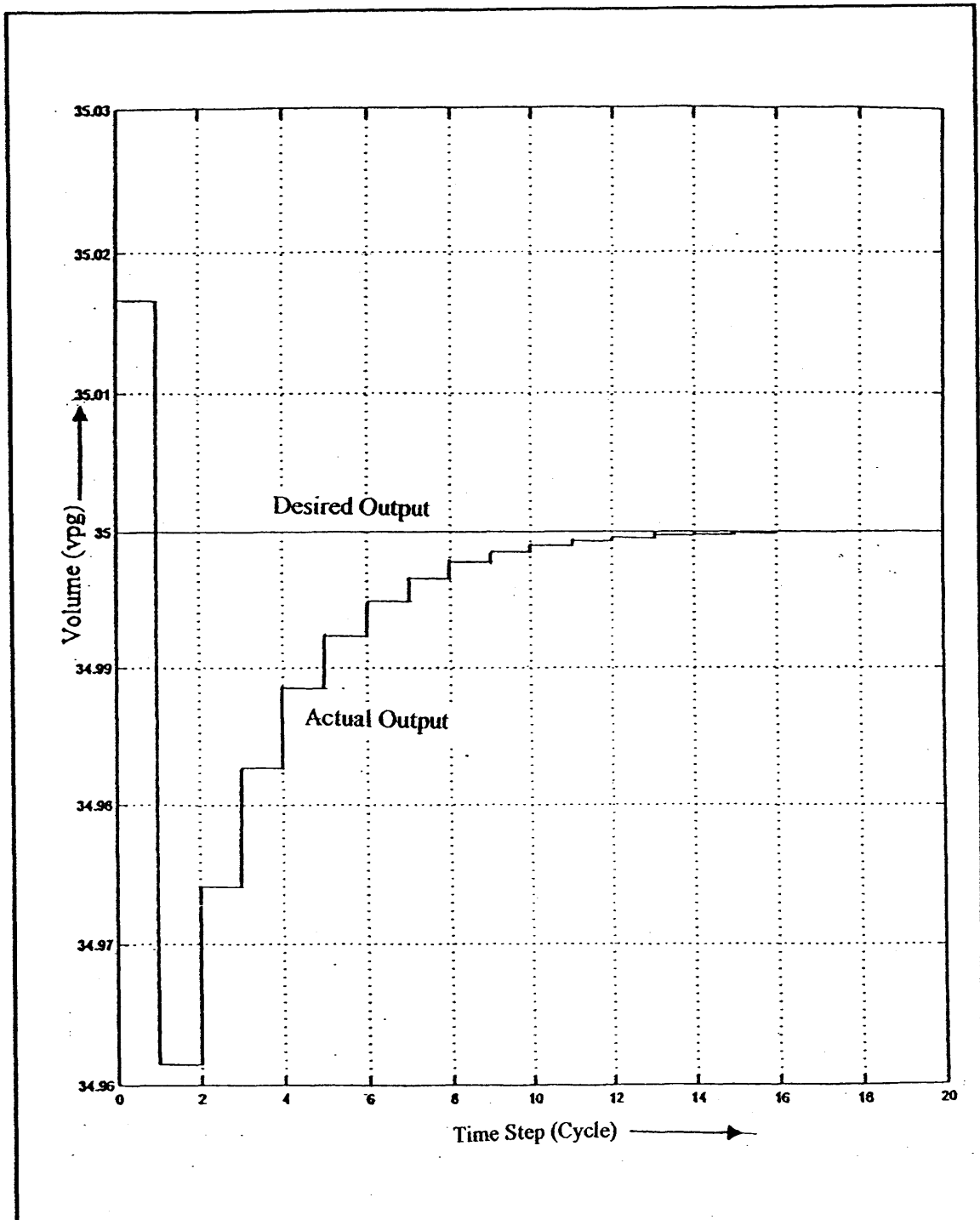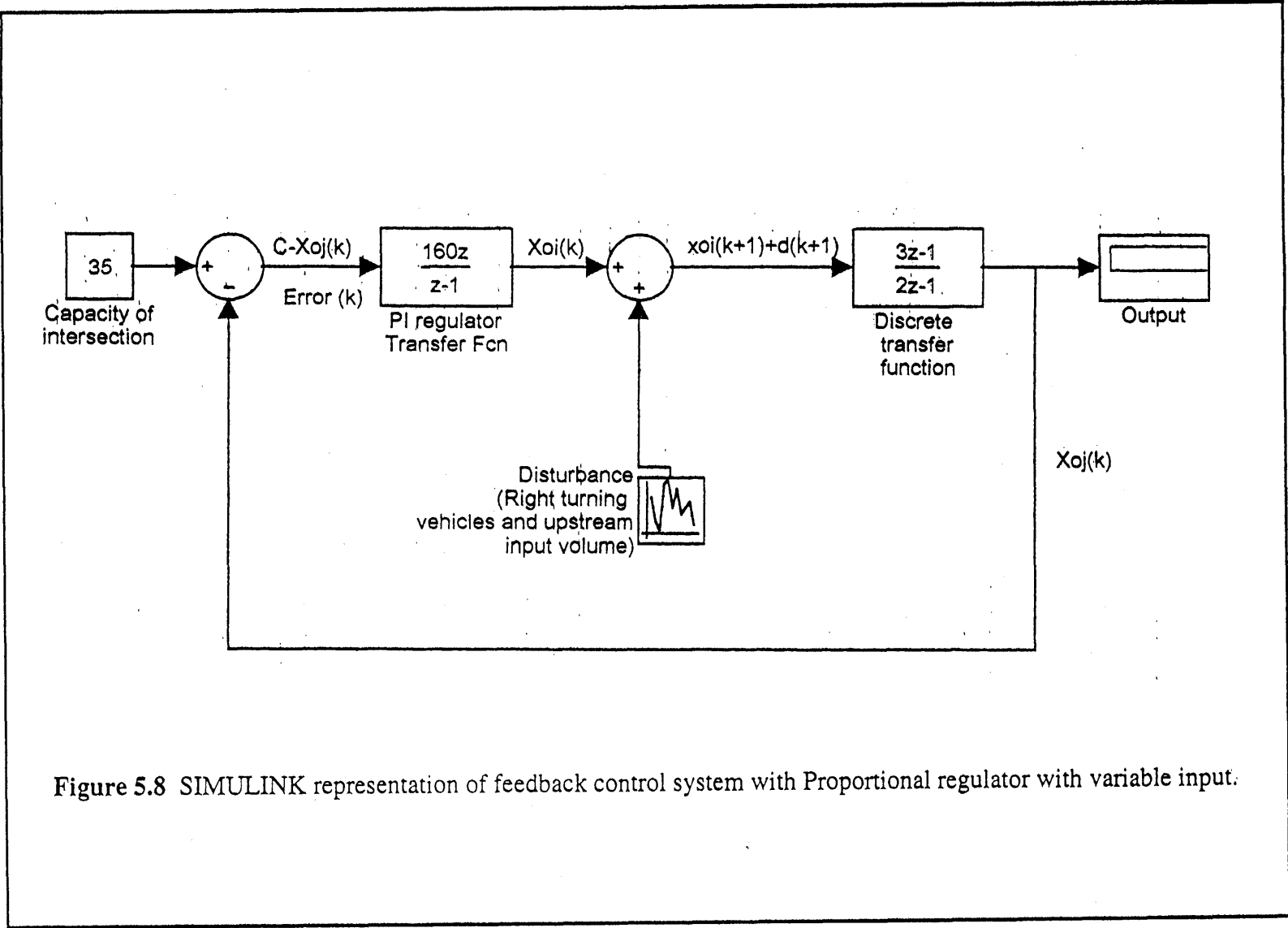
Figure 5.7 Comparison of actual and desired output for a P-controller system with constant input.
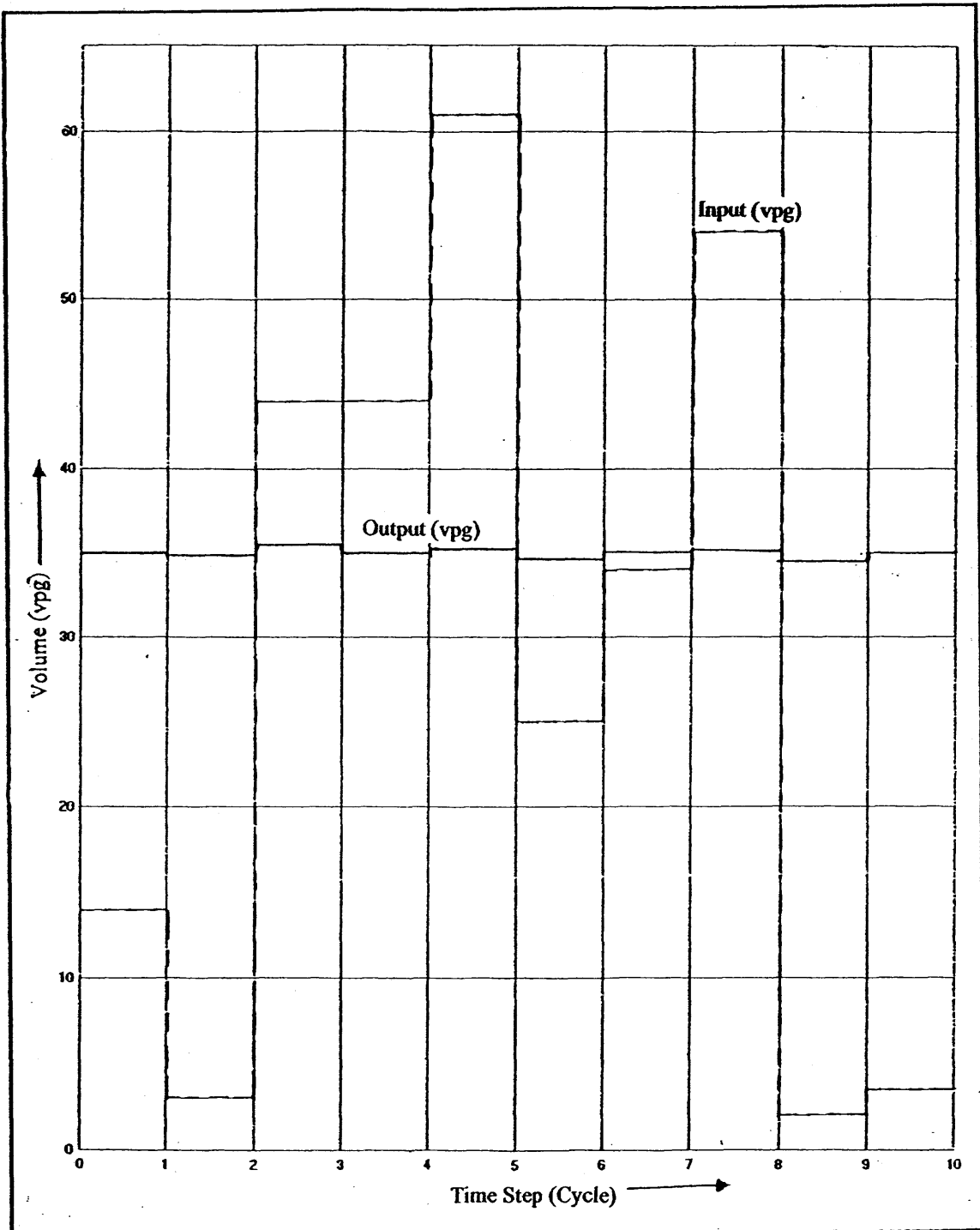
Figure 5.8 SIMULINK representation of feedback control system with Proportional regulator with variable input.

Figure 5.9 Comparison of input and output for a P-controller system with variable input.
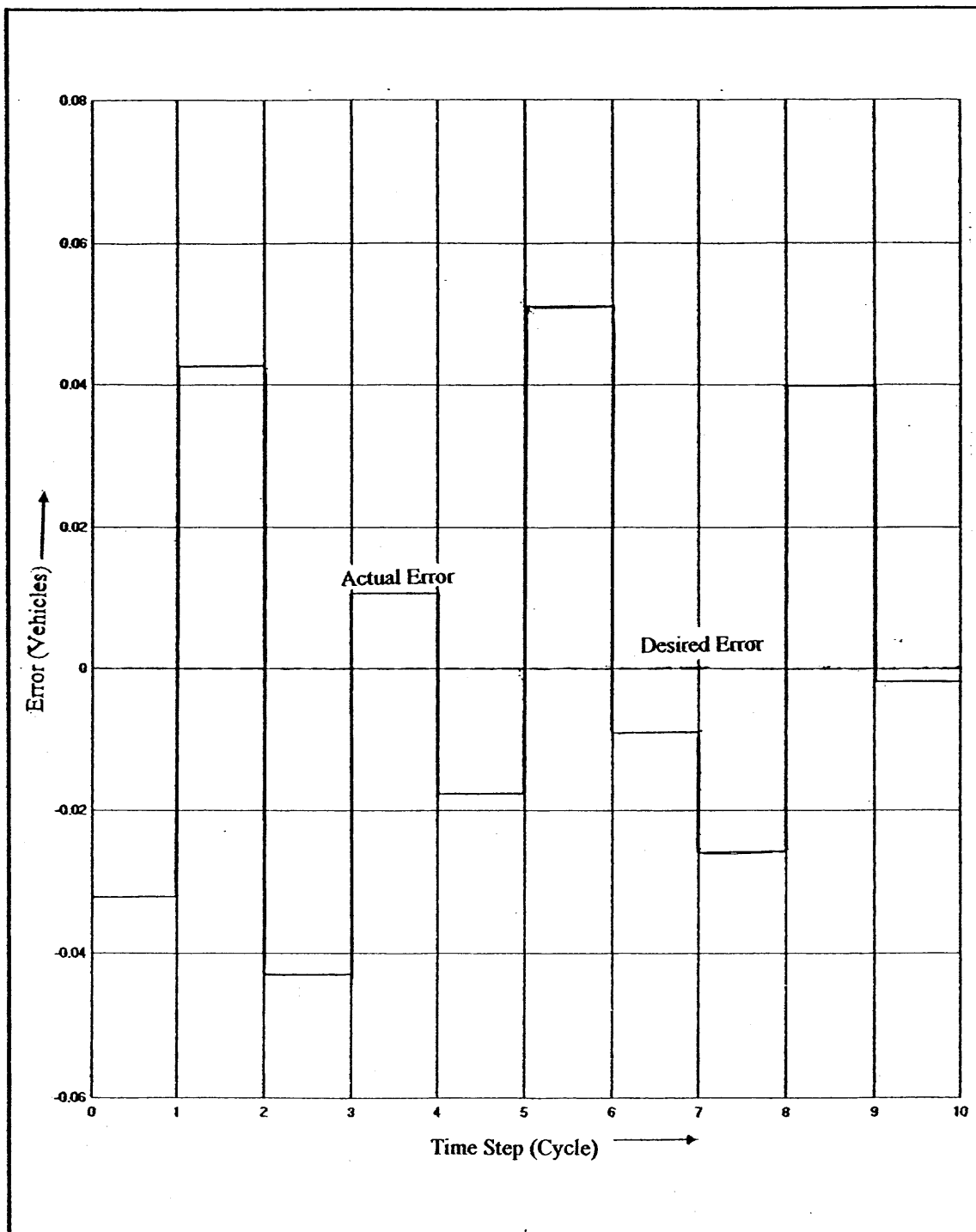
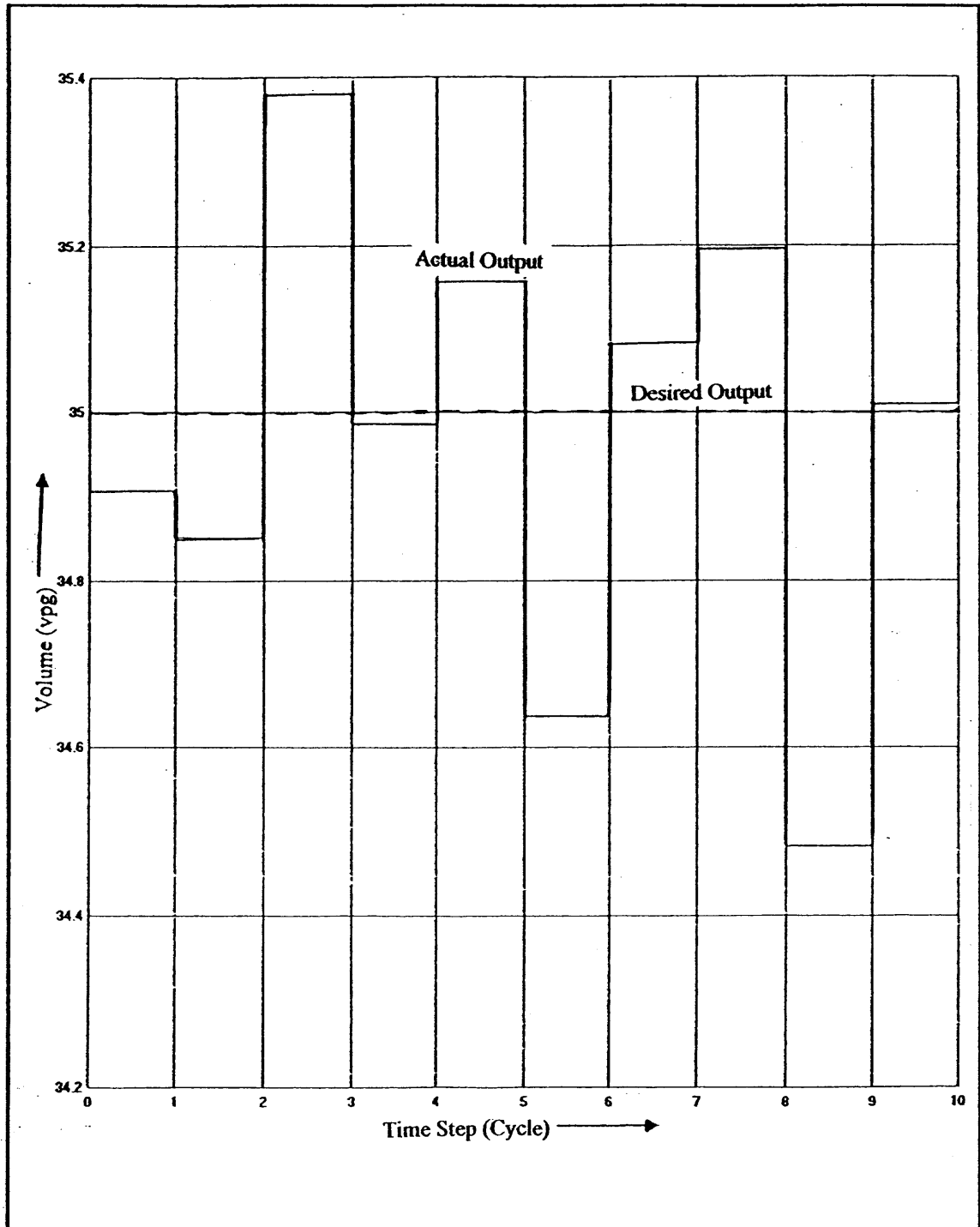**Figure 5.10** Comparison of actual and desired error for a P-controller system with variable input.

Figure 5.11 Comparison of actual and desired output for a P-controller system with variable input.

Case 3. Proportional-Integral Control –Variable Volume

The SIMULINK representation of a control system with a PI-regulator is shown in Figure 5.12. The figure shows the components of the control system in the form of a block diagram. The controller used in this feedback control system is a PI controller with a variable input volume which varies between 0 and 65 vehicles per cycle. Figure 5.13 shows a comparison of input volumes with the output. The figure indicates that the output obtained in this case is close to the desired output than the previous cases of no-control and P-regulator controls, i.e. the system exhibits a higher efficiency in stabilizing the random output to provide a smoother flow of vehicles at the downstream. Figure 5.14 shows the errors of the system with changing input volumes, while Figure 5.15 shows the comparison of the actual and desired error from the system.

Similar to the P-controller analysis, Figures 5.14 and 5.15 show a deviation in error and output from their desired values for PI-controller. However, it is evident, that the PI-controller shows an improvement over the P-controller by tracking the desired values more closely than the P controller.
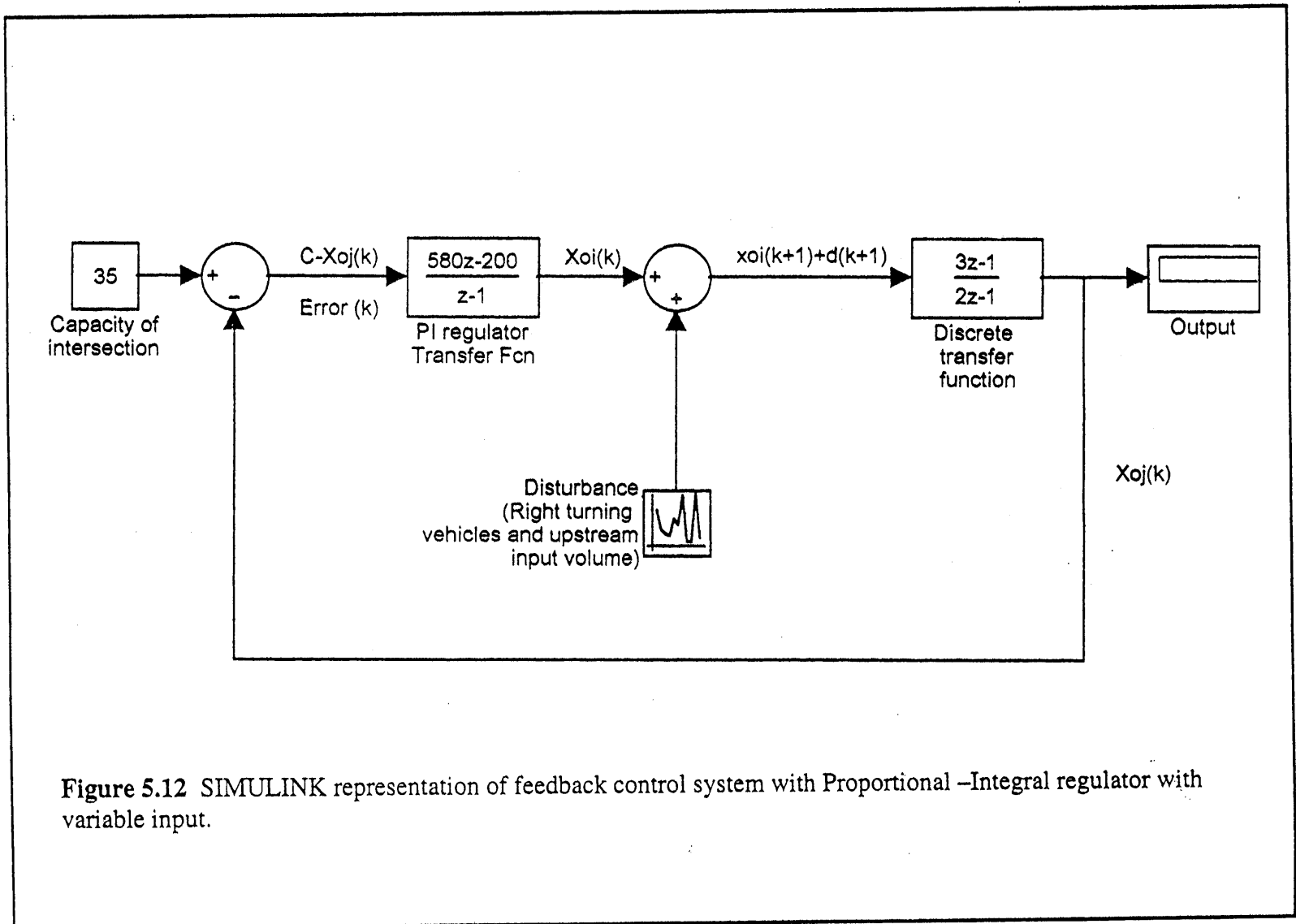
Figure 5.12 SIMULINK representation of feedback control system with Proportional –Integral regulator with variable input.
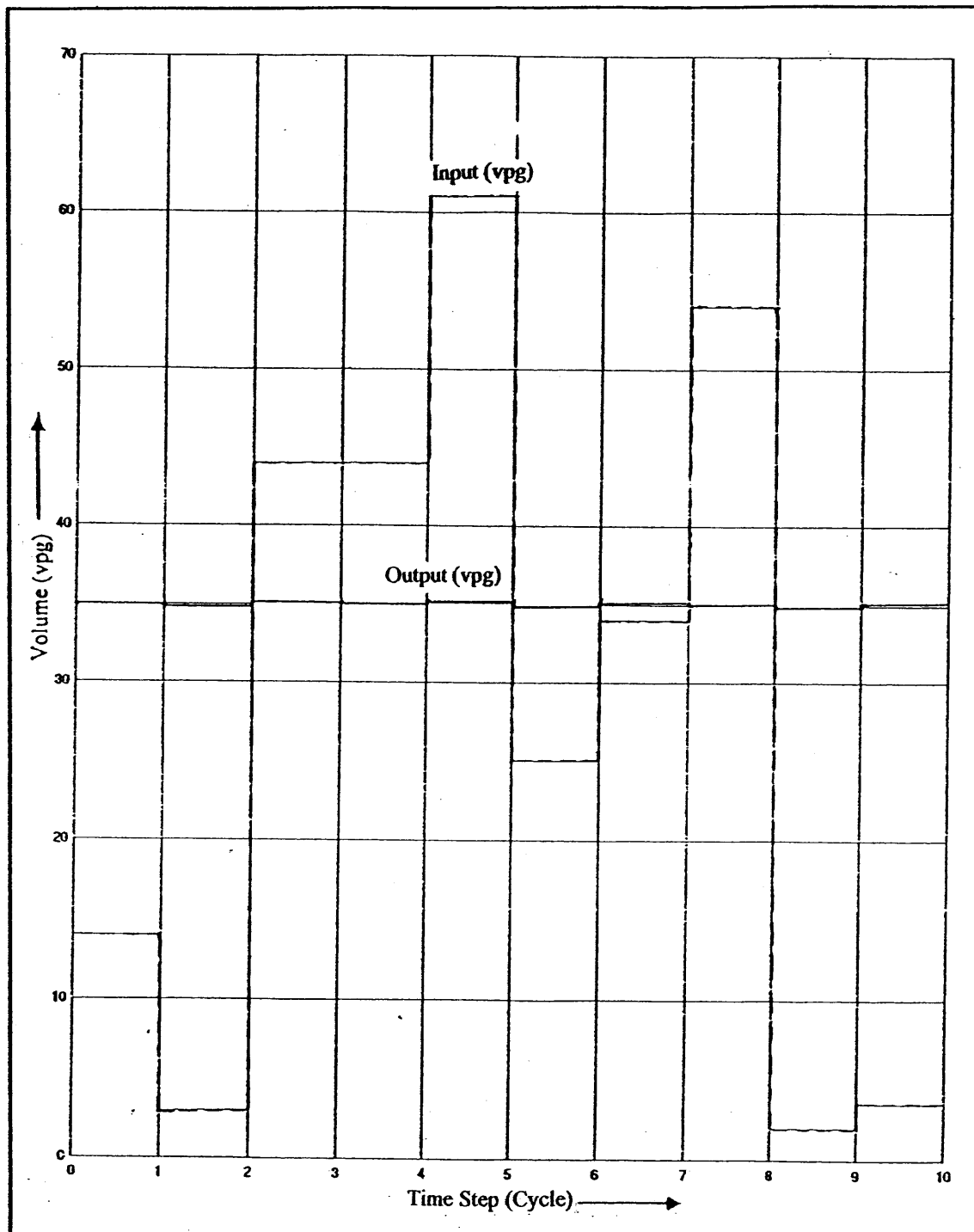
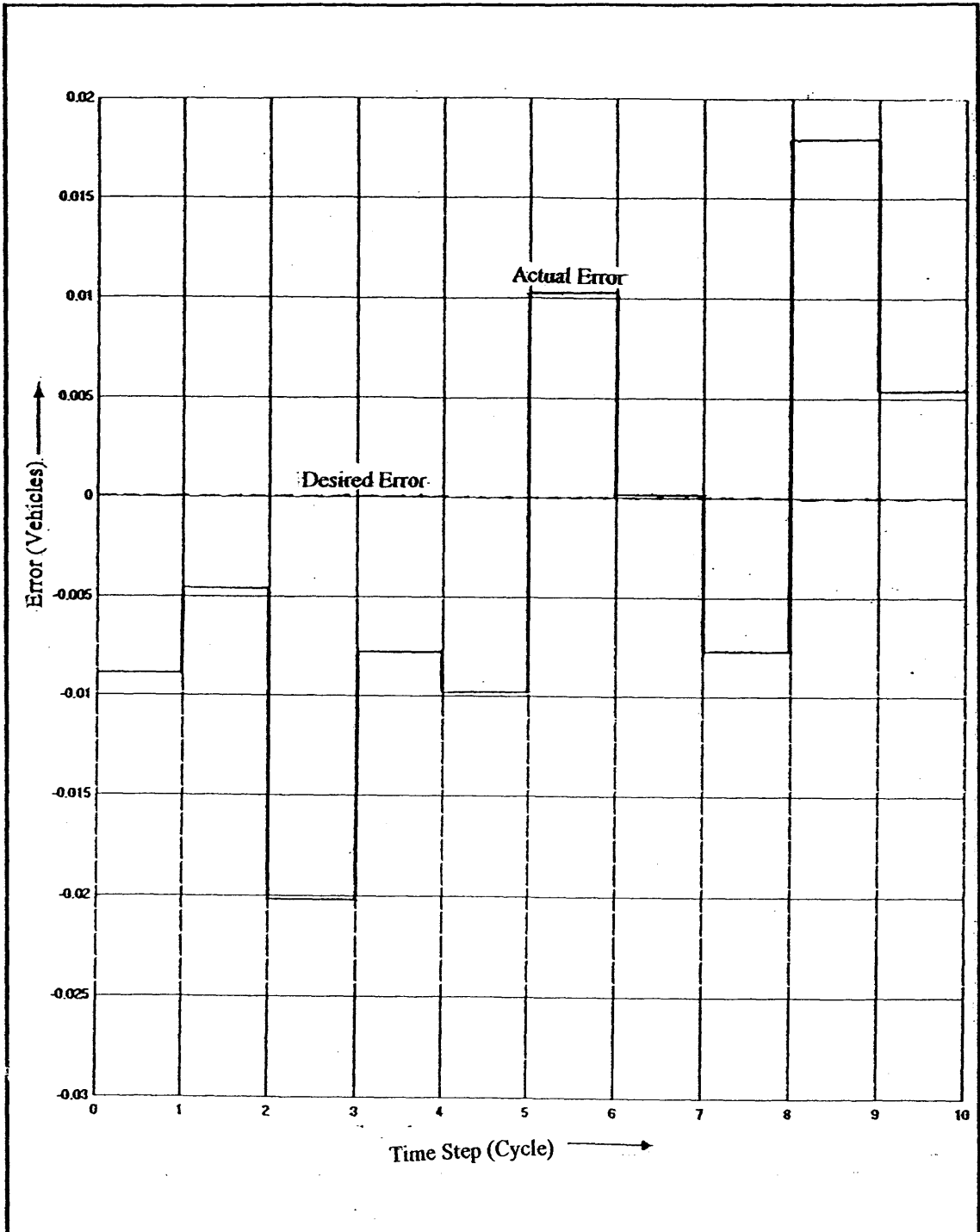**Figure 5.13** Comparison of input and output volumes for a PI-controller system.

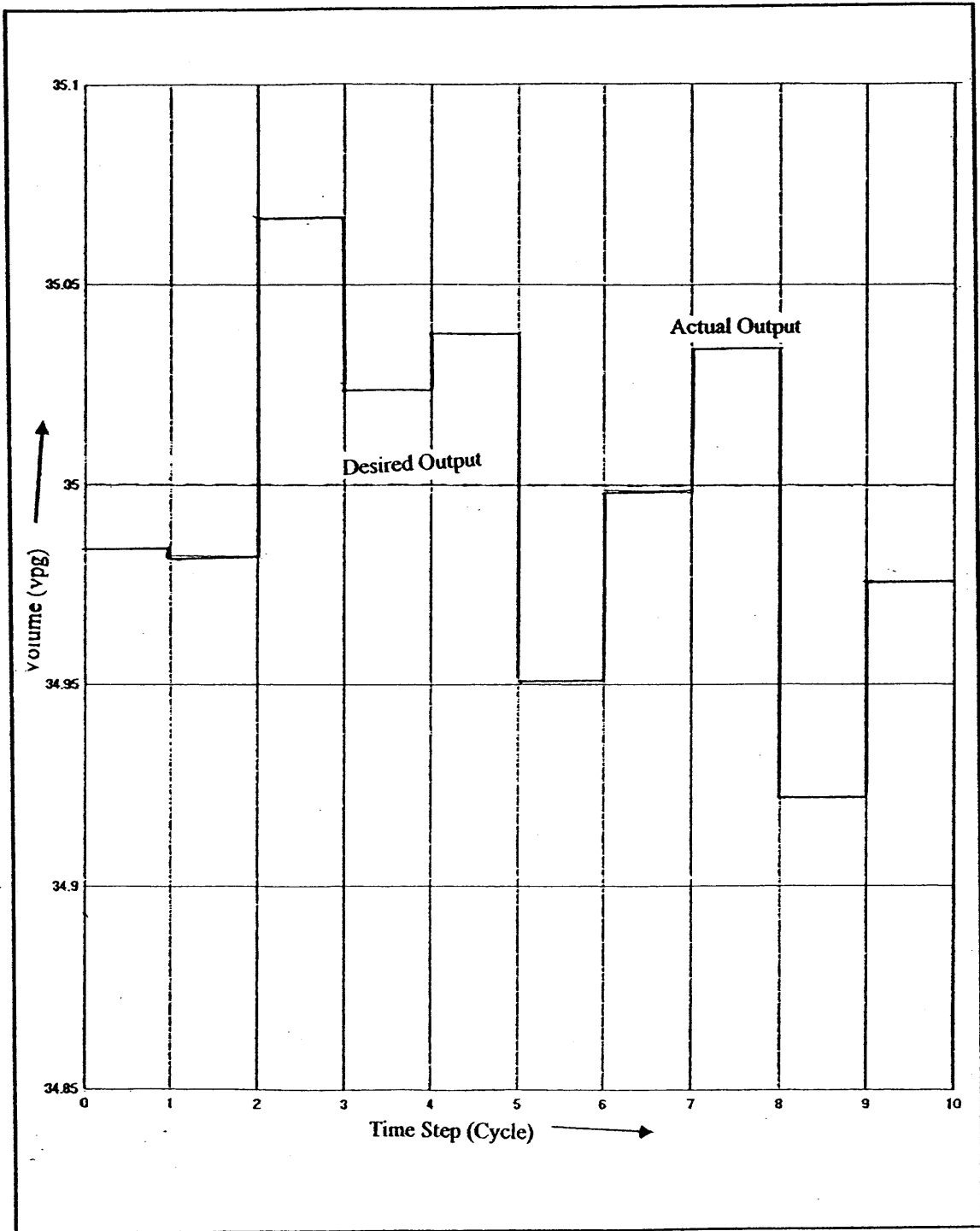**Figure 5.14** Comparison of actual and desired error for a PI-controller system.

Figure 5.15 Comparison of actual and desired output for a PI-controller system.

5.5 Analysis

This section describes the process for selecting the most suitable controller parameter values for each controller (refer to Equations 4.11 and 4.21). The section also describes the best type of controller that should be used to obtain the desired performance from the system. The selection of controller parameter values in this study is based on the criterion of obtaining minimum error and maximum percentage reduction in error to make the performance efficient.

5.5.1 Controller Parameters

The regulator parameter values for both the P and PI controller systems control the errors obtained. For a P controller, the errors are reduced by increasing the value of the regulator parameter K_p. Since it is the property of a proportional controller to increase the overshoot of the control system thus affecting the stability, it is not advisable to consider very high values of K_p. Thus, a tolerance level of ±0.05 is used as a limiting factor for deciding the most suitable value of K_p. The value of K_p, which provides the best combination of considerably lower errors and a high percentage reduction in error, is considered to be the most suitable parameter value for the study system. A similar procedure is used for the PI controller system to obtain the most suitable parameter values of K_p and K_i. However, no optimization routine has been included in this research. Thus, the values of regulator parameters obtained through trial and error method are not optimum.

The percentage reduction in errors for a P controller is shown in Table 5.1. Similar results are obtained for the PI controller. Table 5.1 indicates that the error values keep decreasing with the increase in the value of K_p. An error value of -0.05 is obtained at

$K_p = 160$. The error keeps reducing by further increasing the value of K_p, however, the percentage reduction in error keeps reducing. Thus, it is more beneficial to use a K_p of 160 to achieve a higher reduction in error with a lower instability to the system. A similar analysis for a control system with a PI regulator indicates that a reasonable stability for the control system is achieved at K_p=200 and K_i=380 as indicated in Table 5.2. At these values, there is a significant reduction in the error values for the simulation period, as compared to a P regulator error. Also, the reduction in the error values obtained by further increasing the parameter values is negligible. Beyond this point, very large variation in the parameter values is required to obtain small reduction in errors.

Table 5.1 Variation in Errors with K_p for a P Control System

Kp	Error (Vehicles)*	Percentage reduction in error
10	-0.733	
20	-0.380	48.12
30	-0.259	32.53
40	-0.193	24.50
50	-0.155	19.73
60	-0.13	16.51
70	-0.111	14.11
80	-0.097	12.42
90	-0.087	11.01
100	-0.074	14.06
110	-0.071	9.04
120	-0.065	8.11
130	-0.060	7.72
140	-0.056	7.17
150	-0.052	6.70
160	-0.05	4.58

*Error is defined as the difference in the desired and actual output obtained from the feedback control system at the end of each simulation.

Table 5.2 Variation in Errors with K_p and K_i for a PI Control System

K_p	K_i	Error (Vehicles)
10	30	-0.42
20	50	-0.24
30	80	-0.15
40	70	-0.13
50	80	-0.12
60	140	-0.08
70	160	-0.07
80	170	-0.07
90	190	-0.06
100	220	-0.05
110	250	-0.05
.	.	.
.	.	.
.	.	.
150	320	-0.04
250	390	-0.02

5.5.2 Comparison of P and PI Regulator Performance

To demonstrate the efficiency of controllers in improving the performance of intersection signal control, it is important to draw a comparison between the performance of intersections under the influence of a feedback loop without controllers and with feedback controllers. A further comparison of intersection performance for P and PI controllers provides the controller and controller parameters best suited to serve the given demand for the system under study.

Figures 5.2 and 5.9 show input and output volumes for the no-control and P-regulator cases respectively. A comparison of these figures indicates that the output from the system is fluctuating less and is closer to the desired output as compared to the input volume when controllers are introduced in the feedback system. Figure 5.2 shows a large variation in the output volumes with changing input volumes, but this variation is dampened by the P controller causing the system to fluctuate to a lesser extent. Figure

5.13 shows input and output volumes for the control system with a PI-regulator. A further comparison of Figures 5.9 and 5.13 indicates that the variability in the output is further dampened by introducing a PI controller in the control system. This proves that output of the control system under study can be further stabilized with the help of a PI controller.

Figures 5.3 and 5.10 show actual and desired errors for a no-control case and for a system with a P regulator respectively. The figures indicate that errors for the no-control case range from 0 to 17 units, while the errors for the P controller range from -0.2 to $+0.2$ units. Hence, introducing a Proportional controller in the system makes the system perform close to the desired error of zero. Figure 5.14 shows actual and desired errors for a control system with a PI regulator.

Further comparison of Figures 5.10 and 5.14 indicates that while the errors for the P controller range from -0.2 to $+0.2$, a PI-controller further reduces these errors to a range of -0.05 to $+0.05$. The errors for the same inputs differ widely within the three systems depending on the type of control implementation for the simulation period of 10 time units. Figures 5.3, 5.10 and 5.14 indicate that the error in the system is reduced with the introduction of a P controller, and it is further reduced by the introduction of the PI regulator. Hence, it is evident that the control system helps in controlling the inputs in order to adjust the outputs to desired values, and the PI controller provides the minimum error for the proposed control system for the selected parameter values. It is evident from this discussion that suitable regulator parameters are capable of reducing the errors within the control system by approximately tracking the desired values. However, due to variable inputs, the controllers are unable to accurately track the desired values and reduce the errors of the control system to zero.

Figures 5.4, 5.11 and 5.15 show the actual and desired outputs for the no-control case, system with P regulator and system with PI regulator respectively. A similar comparison of these figures indicates that the actual output follows the desired output more closely in the case of a P controller. Further, comparisons of Figures 5.11 and 5.15 show that the desired output for a PI controller is most closely tracked.

According to the capacity constraint, the output from the system cannot exceed the capacity of the intersection, i.e. the error value should always be positive. The analysis shows that the error values obtained for the two types of control systems attain both positive and negative values. A negative error signifies that input volume is less than the capacity, while a positive error indicates queue formation in case of oversaturation. This explains the likelihood of obtaining both positive and negative error values during the tracking process.

It is evident from the structure of the feedback control system with P and PI regulators, that in the feedback control systems, a regulator is the only element that contributed to tracking the desired output, while in the no-control case, the output is controlled using constraints of capacity and maximum allowable green time. The small errors observed in tracking the desired values by the feedback control systems can be handled by the green time constraints. Thus, the results indicate that a relatively simple feedback loop with the appropriate regulators and regulator parameters can efficiently provide a greater adaptability to signals to cater to the varying demand conditions by reducing the queue formation to a considerable extent as compared to a fixed time signal.

The cycle length is assumed to be constant for a particular simulation run in this research. The cycle length is significant to obtain the capacity of intersections at a given

degree of saturation. A variation in the cycle length leads to a change in the intersection capacity. A sensitivity analysis of the results was also carried out by varying the cycle length and capacity of the intersections. It was observed that the absolute values of errors for same K_p values increased with an increase in the desired capacity. A variation in the desired capacity also affected the error values obtained at the point of steady state. The error values at the steady state increased with an increase in the desired capacity of the system. Thus, it can be concluded that a variation in the cycle length or capacity affects the ability of the system to track the desired value of capacity.

5.6 Comparison of Feedback Control with TRANSYT

As stated earlier, it is out of the scope of this research to field test the adaptive control strategies. However, to test the efficiency of the feedback control systems to handle oversaturated traffic conditions, this research compares the queue lengths of an arterial of two intersections under optimized signal timing obtained from TRANSYT and queue length based on feedback control. Since SCOOT uses a TRANSYT like optimization approach, the simulation of intersections in TRANSYT has been considered to represent the performance of intersection under the SCOOT adaptive traffic control strategy.

A two arterial intersection was simulated using TRANSYT-7F (Release 9) and MATLAB. The assumed operating conditions included a constant input flow of 65 vpg (2340 vph), a saturation flow rate of 1800 vphg, a degree of saturation of 1.0 and a cycle length of 100 seconds. The simulation of the arterial was carried out for 16 cycle lengths in both TRANSYT and MATLAB. The system performance in TRANSYT is measured in terms of total travel time, delays, stops, queues and level of service. The MOE

considered in this research to compare the results of the two models was the queue length. A constant volume of 2340 vph was fed into the arterial which represented a highly oversaturated volume condition. At the end of 16 time periods, i.e. 16 cycle lengths (27 min), the queue generated by the optimized performance under TRANSYT was compared with the queue at the end of 16 time units generated by a P controller.

The results obtained by TRANSYT indicated that at the end of simulation under optimized conditions, the average queue length in downstream link was 9 vehicles/link. Since the queue holding capacity of the link was specified as 0 vehicles/lane for both simulation models, TRANSYT generated a spillback. The MATLAB results for a P controller feedback system under constant input volume conditions (refer to Figure 5.6) indicated that queue at the end of 16 cycles was zero. By operating the arterial under feedback control, a reduction of 12.5% in the cumulative queue lengths was obtained for on the main street during the simulation period. This can be explained by the fact that since the capacity of the downstream intersection is fixed to 35vpg, the feedback control system meters the input volumes from upstream. Thus, depending on the queue lengths at the end of each cycle, the control system updates the amount of inflow allowed from upstream intersection to the downstream intersection. The inflow from upstream, or demand of the downstream intersection to operate the arterial under capacity conditions is updated every cycle.

Comparing the results of TRANSYT and MATLAB indicates that feedback control systems possess the ability of timely updating varying demand conditions. This property of feedback control systems can be used to overcome some of the deficiencies in

the estimation procedures of existing adaptive traffic control strategies such as SCOOT, SCATS and OPAC.

5.6 Summary

The results obtained by MATLAB prove that a well designed feedback control is an efficient tool for obtaining better queue management in arterials. The same theory can be applied to larger networks for a better performance of individual intersections and networks. As described in Chapter 3, one of the primary reasons for a sub-optimal performance of SCOOT, SCATS and OPAC has been the feed-forward operation of these strategies. The feed-forward mechanism of data flow does not allow the variable quantities to be updated in time. The conceptual framework proposed in this research is an attempt to overcome the deficiencies introduced in adaptive control systems due to a feed-forward operation by utilizing the robustness of a feedback control systems. The results indicate that feedback control systems, with well designed controllers exhibit efficient queue handling capabilities during oversaturation. This feature of feedback control leads to a stable system providing nearly desired performance after a few iterations.

Under constant volume conditions, feedback control systems work with absolute accuracy by gradually brining the errors in the system to an absolute zero within a few time periods. The performance of these systems in case of variable volume conditions is also considerable. Hence, the feed-forward mechanism of current adaptive traffic control systems can be improved by providing timely updates to demand and queues through a

feedback control system. The updates reduce inaccuracies associated in the estimation processes which are currently in use by the adaptive traffic control systems.

CHAPTER 6

CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary

Most of the existing adaptive traffic control strategies follow a feed-forward data flow mechanism. Although these strategies use elaborate algorithms to address the problems of oversaturated conditions and queue formation, a number of shortcomings are introduced as a result of the feed-forward operation. These shortcomings are due to the lack of a corrective mechanism to adjust the signal control system to reflect dynamic traffic flow patterns. This research provided an approach to use the properties of feedback control to overcome some of the deficiencies of feed-forward control systems.

This research proposed a conceptual framework to operate existing adaptive traffic control systems in a feedback environment and enhance the performance of these systems through a better estimation of variable demands and the efficient handling of queues. The proposed framework was applied to three well known adaptive traffic control strategies: SCOOT, SCATS, and OPAC. A detailed study of adaptive control procedures of the three systems revealed an inadequacy in demand estimation and queue handling procedures introduced due to their feed-forward operation. The research applied the proposed framework to the operation of the three systems and predicted an enhancement in their performance through the use of concepts of feedback control. It is out of the scope of this research to test the performance of SCOOT, SCATS and OPAC for their performance in a feedback control environment. However, the efficiency of feedback control systems can be tested by applying the concept to a hypothetical arterial.

Hence, a feedback control system was designed for an arterial of two intersections. Two types of controllers, Proportional and Proportional-Integral controller were designed for the arterial signal system.

The performance of the arterial under variable input volumes was tested using MATLAB, a powerful tool for the analysis of control systems. The performance of the arterial was assessed by first operating the arterial under a feedback loop without the use of controllers and then by using a proportional controller for a given volume condition. It was found that the proportional controller improved the performance by smoothing the output from the downstream intersection and by controlling a highly variable input volume at the upstream intersection. The results also indicate that feedback control produces better queue management. The same arterial was tested using a PI controller and its performance was compared with the no-control and P controller cases. It was found that the system performed best using a PI controller. Further testing of the system under constant volume conditions indicated that zero errors, i.e. output equal to capacity for oversaturated conditions, can be obtained by providing constant input to the feedback control system. The performance of the same arterial under an optimal signal timing plan was tested in TRANSYT. The system inputs used were the same as in MATLAB. The queue handling efficiency for the two simulation models was compared.

The study proved the efficiency of feedback control systems in enhancing traffic signal control systems. The study also proved the ability of a PI controller to best handle the given volume and capacity conditions for the study arterial system. The results indicated that an adaptive traffic control system if operated under a feedback control environment can efficiently handle queues and accurately assess the variable demands for

subsequent cycles. Thus, the proposed framework has a potential benefit in improving the performance of existing adaptive traffic control systems.

6.2 Conclusions

The following conclusions can be drawn from this study regarding the ability of feedback control enhance existing adaptive traffic control systems:

- The corrective properties of feedback control can be used in adaptive traffic control systems to timely update the dynamically changing parameters to obtain desired performance, specifically during oversaturated conditions.

- The conceptual framework integrating adaptive traffic control strategies with concepts of feedback control has the potential to enhance the current adaptive control systems by providing means of better estimation of demand and evaluation of performance.

- A simple feedback loop with appropriate regulators and regulator parameters can efficiently provide adaptability to traffic signals to handle both constant, as well as varying demand conditions.

- Controllers play a key role in regulating the control variables as desired. A feedback loop without controllers cannot be recommended as a tool for queue management.

- A PI regulator with appropriate regulator parameter values provides low errors with acceptable stability. Thus, it is best suited for arterials. A PID controller may further improve the system performance, but has not been tested in this study.

- Properly designed feedback control systems can also act as a standalone strategy to obtain better queue management within individual intersections and arterials. These systems can be designed and modified to meet specific requirements of traffic control, such as the calculation of appropriate signal timing plans.

6.3 Recommendations for Future Research

This research provided a conceptual framework for introducing the concept of feedback control into the operation of adaptive traffic control systems. The study can be further strengthened by applying a feedback control system into the algorithm of the three adaptive control systems considered in this research and performing appropriate field tests. The testing can further be extended to a network of intersections. This research followed a trial-and-error technique to obtain at the most appropriate values of regulator parameters. However, this technique does not provide most accurate solutions and it is time consuming. The research can be further extended to include an optimization routine to obtain accurate values of regulator parameters. Also, the scope of this research has been limited to testing the arterial system under P and PI regulators. Therefore, it is recommended that the study should be further extended to explore the efficiency of PID controllers, which proved to be efficient in other industrial processes.

APPENDIX A

DEFINITIONS

Appendix A consists of the definitions and specifications of the transient-response characteristics of a control system.

A.1 Definitions

This section provides the definitions and generally used specifications of the transient-response characteristics of control system to a unit-step input *(Ogata, 1990)*. Figure A.1 shows the graphical representation of the unit-step response curve indicating the transient-response characteristics.

1. Delay Time (t_d)

The delay time is the time required for the response to reach half the final value the very first time.

2. Rise Time (t_r)

The rise time is the time required for the response to rise from 10% to 90%, 5% to 95%, or 0% to 100% of its final value.

3. Peak Time (t_p)

The peak time is the time required for the response to reach the first peak of the overshoot.

4. Maximum (Percent) Overshoot (M_p)

The maximum overshoot is the maximum peak value of the response curve measured from unity.

5. Settling Time (t_s)

The settling time is the time required for the response curve to reach and stay within a range about final value of size specified by absolute percentage of the final value (usually 2% or 5%).
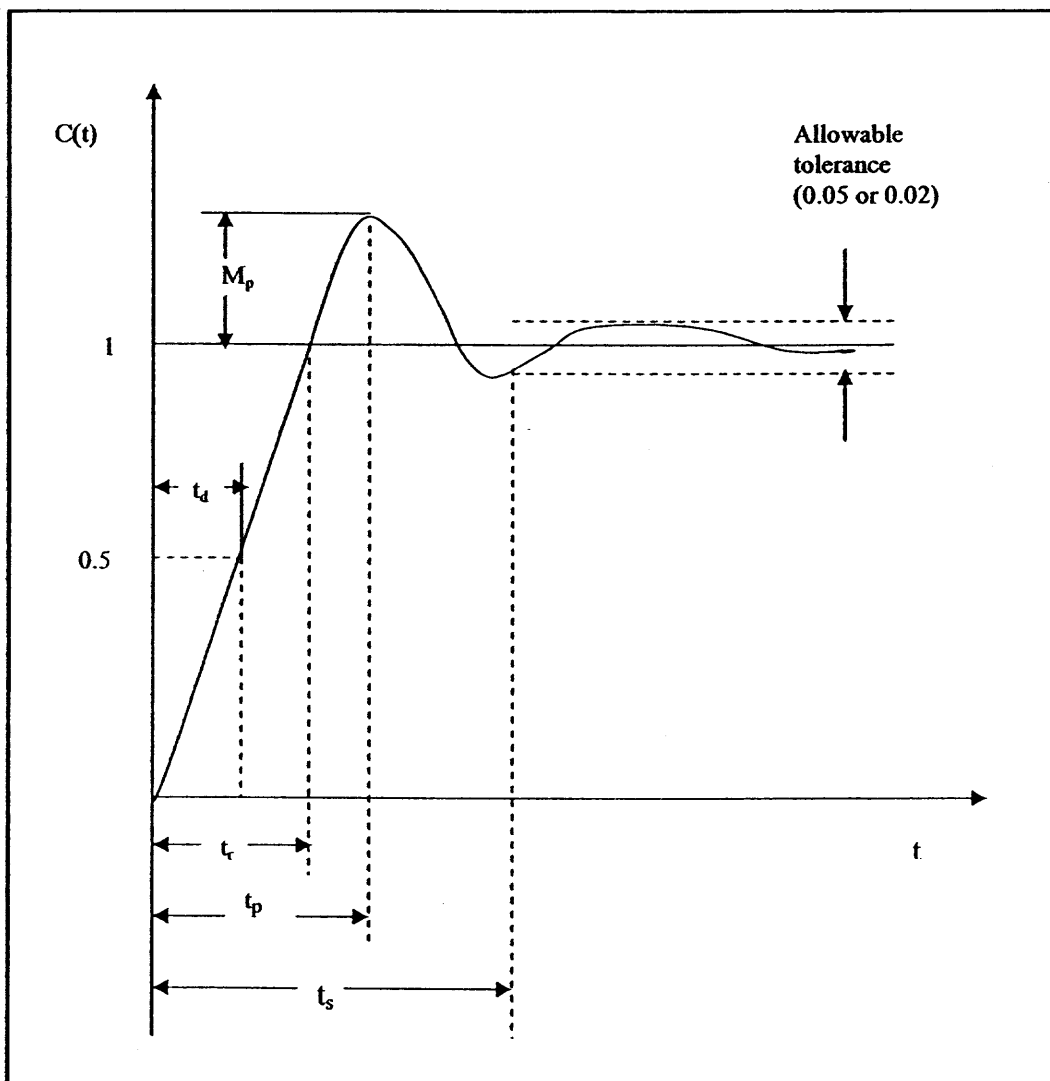
Figure A.1 Unit-step response curve showing t_d, t_r, t_p, M_p, and t_s.

APPENDIX B

CHARACTERISTICS OF CONTROLLERS

Appendix B consists of the characteristics of Proportional (P), Integral (I) and Derivative (D) controllers. The section includes the properties of the three types of controllers in terms of the transient properties of systems.

B.1 Characteristics of P, I and D Controllers

A proportional controller (K_p) has the effect of reducing rise time, but it can never eliminate the steady-state error. An integral controller (K_i) has an effect of eliminating the steady-state error. However, it worsens the transient response of the system. A derivative control (K_d) has an ability to increase the stability of the system, reducing the overshoot and improving the transient response. These effects can be summarized as shown in Table B.1.

Table B.1 Characteristics of P, I and D Controllers

Controller Response	Rise Time	Overshoot	Settling Time	Steady-state Error
K_p	Decrease	Increase	Small change	Decrease
K_i	Decrease	Increase	Increase	Eliminate
K_d	Small change	Decrease	Decrease	Small Change

APPENDIX C

Mass-Spring-Damper Control Problem

Appendix C consists of an example demonstrating the application of Proportional (P), Proportional-Integral (PI) and Proportional-Integral-Derivative (PID) controllers on a simple mass spring and damper problem. The example shows the effect of the three types of controllers on the system.

EXAMPLE

The effect of application of P, PI and PID regulators on a system can be studied with the help of a simple mass-spring-damper system *(Refer Control Tutorials for MATLAB)*.
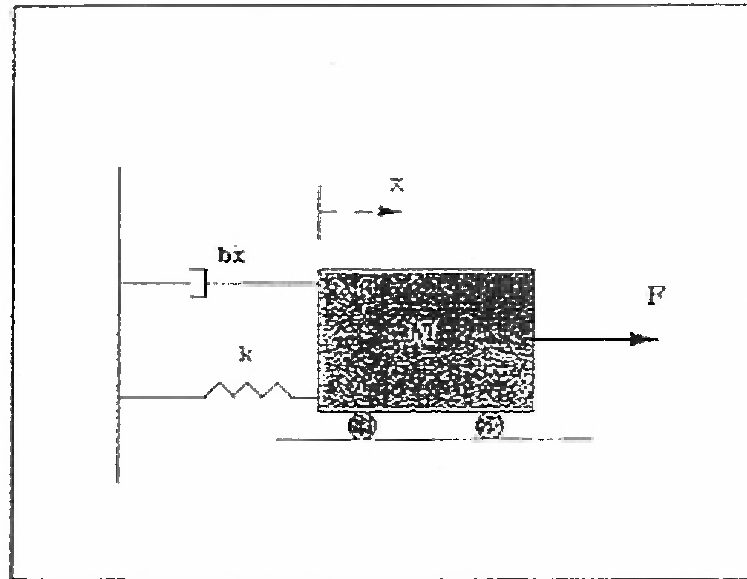

Figure C.1 Simple mass, spring and damper system.

The solution begins with obtaining the modeling equation for the mass spring system as shown in Figure C.1. The figure shows a mass M attached to a fixed support through a spring with a spring coefficient k. An external force F is applied to the mass which causes the mass to displace by a distance x. This example analyses the effect of P, PI and PID controllers in bringing the oscillations of the system to zero. The goal of this problem is to show how each of Kp, Ki and Kd contributes to obtain a fast rise time, minimum overshoot, and no steady-state error.

The modeling equation of this system can be given as follows:

$$M\ddot{x} + b\dot{x} + kx = F \tag{C.1}$$

where M represents the mass, F represents the external force applied to the system, k denotes the spring stiffness coefficient, b represents the damping coefficient and x represents the displacement.

In order to solve this problem using control theory, it is required to obtain the transfer functions between input and output. Input in this case is the external force F acting upon the mass and the output is the displacement obtained as a result of the application of force. The transfer functions are obtained using Laplace transform.

Taking the Laplace transform of the modeling equation (C.1)

$$Ms^2 X(s) + bsX(s) + kX(s) = F(s) \tag{C.2}$$

The transfer function between the displacement X(s) and the input F(s) then becomes

$$\frac{X(s)}{F(s)} = \frac{1}{Ms^2 + bs + k} \tag{C.3}$$

The similar transfer can be obtained using difference equations instead of differential equations as the modeling equation of the system. Also, differential equation can be discretized into a difference equation with boundary conditions. The z-transform can be used to obtain the transfer function between the displacement and the input instead of the Laplace transform, which is typically used with differential equations. A Laplace transform can also be converted to a z-transform.

Consider values of mass M as 1kg, b as 10 N.s/m, k as 20 N/m and force F(s) = 1 N. Substituting these values into the above transfer function, the following transfer function can be obtained:

$$\frac{X(s)}{F(s)} = \frac{1}{s^2 + 10s + 20}$$ (C.4)

The gain of the plant transfer function is 1/20, so 0.05 is the final value of the output to a unit step input. This corresponds to the steady-state error of 0.95 which is quite large. It can also be seen from Figure A.2 that the rise time is about one second, and the settling time is about 1.5 seconds. Now, it is desired to design a controller that will reduce the rise time, reduce the settling time, and eliminates the steady-state error.
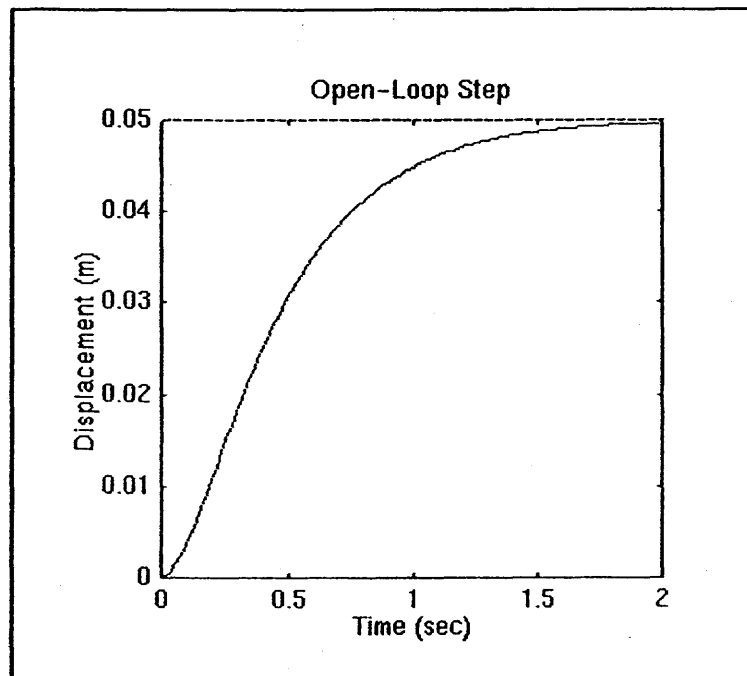

Figure C.2 Displacement with time.

Case 1 Proportional Control

From Table B.1, we see that the proportional controller (K_p) reduces the rise time, increases the overshoot, and reduces the steady-state error. The closed-loop transfer function of the above system with a proportional controller is:

$$\frac{X(s)}{F(s)} = \frac{K_p}{s^2 + 10s + (20 + K_p)}$$

(C.5)

The value K_p is found by trial and error method with the help of properties of P, I and D controllers. Let the proportional gain (K_p) be 300. Figure C.3 shows the effect of the proportional controller on the displacement of the mass M, with respect to time. The plot indicates that the proportional controller reduced both the rise time and the steady-state error, increased the overshoot, and decreased the settling time by small amount.
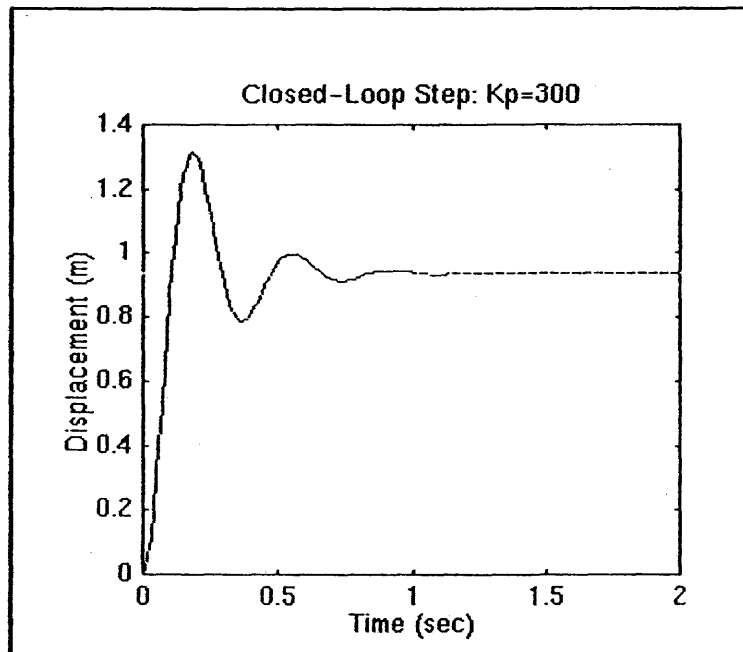

Figure C.3 Effect of Proportional control.

Case 2 Proportional-Derivative control

From Table B-1, it is known that the derivative controller (K_d) reduces both the overshoot and the settling time. The closed-loop transfer function of the given system with a PD controller is:

$$\frac{X(s)}{F(s)} = \frac{K_D s + K_p}{s^2 + (10 + K_D)s + (20 + K_p)} \tag{C.6}$$

Assuming the value of K_p equal to 300 and K_d equal to 10, the plot of displacement against time is obtained. Figure C.4 shows that the derivative controller reduced both the overshoot and the settling time, and had small effect on the rise time and the steady-state error.
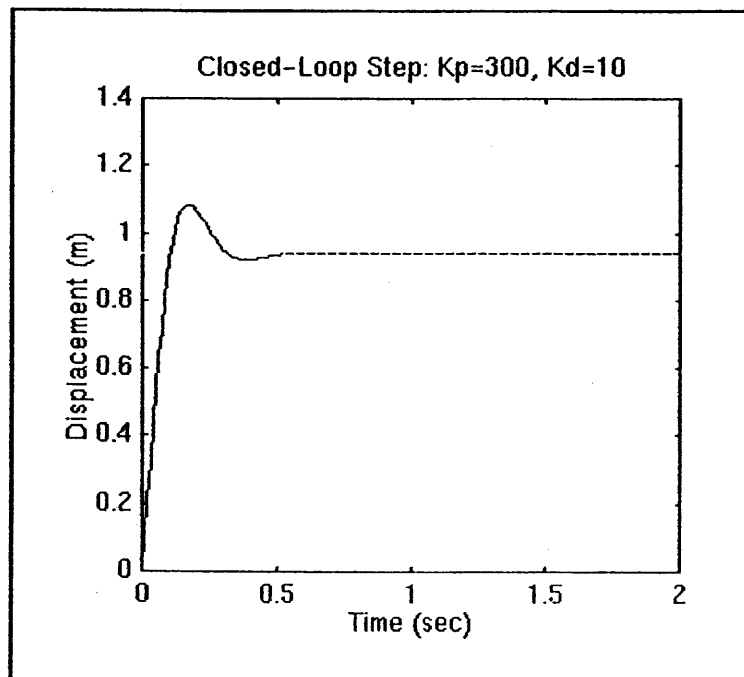

Figure C.4 Effect of Proportional-Derivative control.

Case 3 Proportional-Integral control

An integral controller (K_i) decreases the rise time, increases both the overshoot and the settling time, and eliminates the steady-state error. For the given system, the closed-loop transfer function with a PI control is:

$$\frac{X(s)}{F(s)} = \frac{K_p s + K_I}{s^2 + 10s^2 + (20 + K_p)s + K_I}$$

(C.7)

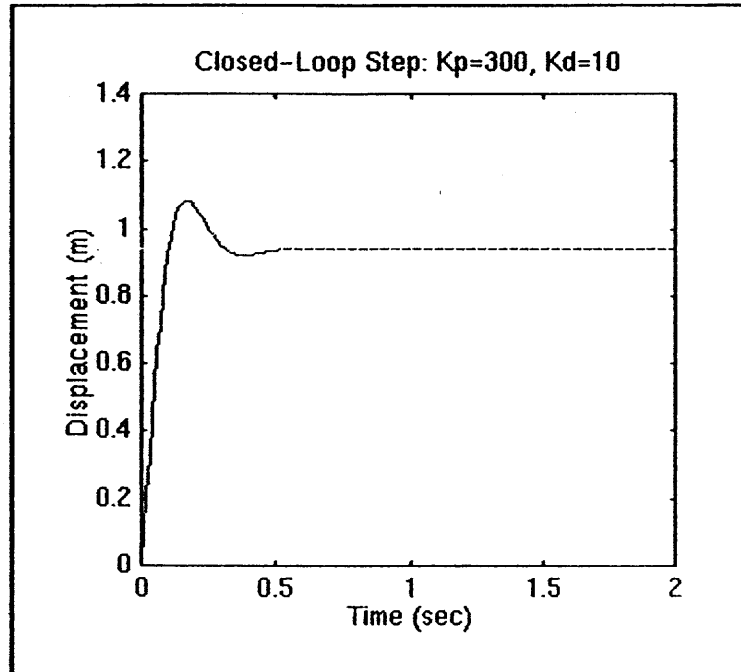In this case, K_p is reduced to 30 and K_i is taken equal to 70.

Figure C.5 Effect of Proportional-Integral control.

The proportional gain (K_p) has been reduced in this case because the integral controller also reduces the rise time and increases the overshoot as the proportional controller does (double effect). The response represented in Figure C.5 shows that the integral controller eliminated the steady-state error.

Case 4 Proportional-Integral-Derivative Control

The closed-loop transfer function of the given system with a PID controller is:

$$\frac{X(s)}{F(s)} = \frac{K_D s^2 + K_p s + K_I}{s^3 + (10 + K_D)s^2 + (20 + K_p)s + K_I}$$

(C.8)

Selection of regulator parameters through trial and error indicates that the gains K_p=350, K_i=300, and K_d=50 provided the desired response. The response of a PID control is shown in Figure C.6. The figure indicates that under the effect of PID control, the system is obtained with no overshoot, fast rise time, and no steady-state error.
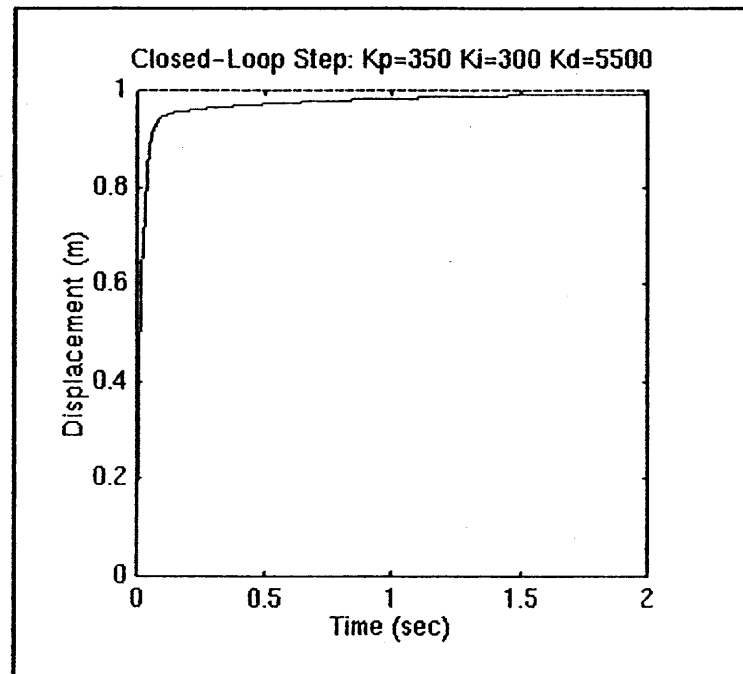

Figure C.6 Effect of Proportional-Integral-Derivative Control.

REFERENCES

Abu-Lebdeh, G. and R.F. Benekohal. Micro-Genetic Algorithms in Intelligent Traffic Signal Control. *Fifth International Conference on Applications of Advanced Technologies in Transportation Engineering*, 1998, pp. 288-295.

Banos, J.C.M. and M. A. Papageorgiou. Linear Programming Approach to Large-Scale Optimal Control Problems, *IEEE Transactions on Automatic Control*, Vol. 40, Issue 5, 1995, pp. 971-977.

Belanger, P. R. (1995). Control Engineering-A Modern Approach. Fort Worth: Saunders College Publishing,.

Bogart, Theodore F. Jr. (1982). Laplace Transforms and Control System Theory for Technology, New York: John Wiley & Sons.

Bretherton D., Wood K., Baker, K. and B. Radia. Congestion and Incident Management using the SCOOT UTC System. *Tenth International Conference on Road Transport Information and Control*, Conference Publication, No. 472, 2000, pp. 96 -100.

Carnegie Mellon. *Control Tutorials for MATLAB*. http://www.engin.umich.edu/group/ctm/PID/PID.html (12 Feb 2002).

David I. Robertson. Handling Congestion with SCOOT. *Journal of Traffic Engineering and Control*, Vol. 28, No. 4, 1987, pp. 228-230.

David I. Robertson. Research on the TRANSYT and SCOOT Methods of Signal Coordination. *Journal of Institute of Transportation Engineers*, Vol. 56, No. 1, Jan 1986, pp. 36-40.

DeRusso, P. M., Roy, R. J., & Close, M. C. (1965). State Variables for Engineers, New York: John Wiley & Sons, Inc.

Ellis, George. (1991). Control System Design Guide (2nd ed.). San Diego: Academic Press.

Franklin, G. F., Powel, J.D. & Emami-Naeini A. (1994). Feedback Control of Dynamic Systems. (3rd ed.). Reading, Mass: Addison-Wesley Publishing Company.

Friedland, Bernard. Control System Design (1986). An Introduction to State-Space Methods. New York: McGraw-Hill Book Company.

Gartner, N. H., Tarnoff, P. J., and C.M. Andrews. Evaluation of Optimized Policies for Adaptive Strategy. *Transportation Research Record 1324*, 1991, pp. 105-114.

Gartner N.H., Pooran, F.J. and C.M. Andrews. Implementation of the OPAC Adaptive Control Strategy in a Traffic Signal Network. *Proceedings of Intelligent Transportation Systems*, 2001, IEEE, pp. 195 -200.

Gartner, N. H. OPAC. A Demand Responsive Strategy for Traffic Signal Control. *Transportation Research Record,* No. 906, 1985, pp. 75-81.

Gartner, N. H. OPAC: Demand-Responsive Traffic Signal Control Research. Transportation Res.-A, Vol. 19A, No. 5/6, 1985, pp. 373-376.

Hadj-Salem H. and Blosseville, J.M. and M. Papageorgiou. ALINEA: A Local Feedback Control Law for On-Ramp Metering: A Real-Life Study, *Third International Conference on Road Traffic Control*, 1990, pp. 194 –198.

Hamad and Kikuchi. Developing a Measure of Traffic Congestion: A Fuzzy Inference Approach. *Proceedings of 81st Transportation Research Board Annual Meeting, Washington, DC., Preprint CD-ROM*, 2002.

Head, K.L., Mirchandani, P.B. and D. Sheppard. Hierarchical Framework for Real-Time Traffic Control. *Transportation Research Record*, No. 1360, 1992, pp. 82-88.

Hua J. and A. Faghri. Development of Neural Signal Control System- Toward Intelligent Traffic Signal Control. *Transportation Research Record,* No. 1497, 1995, pp. 53-61.

Hunt, P.B., Robertson, D.I. and M.C. Royle. The SCOOT On-Line Traffic Signal Optimization Technique. *International Conference on Road Traffic Signaling*, Vol. 207, 1982, pp. 59-62.

Jury, E. I. (1964). Theory and Application of the z-Transform Method. New York: John Wiley & Sons, Inc.

Khatib K. Zaher and Geoffrey A. Judd. Control Strategy for Oversaturated Signalized Intersections. *Proceedings of 80th Transportation Research Board Annual Meeting, Washington, DC,* Jan 2001.

Kotsialos, A., Papageorgiou M., Mangeas, M. and H. Haj-Salem. Coordinated and Integrated Control of Motorway Networks via Non-Linear Optimal Control. *Transportation Research-C,* Vol. 10, 2002, pp. 2002.

Li, H. and Prevedouros P. Traffic Adaptive Control for Over-Saturated Isolated Intersections. *Presented at Annual Meeting of the Transportation Research Board,* 2002.

Liu, H. and M. Kuwahara. An Approach on Network Traffic Signal Control under the Real-Time and Oversaturated Flow. *Proceedings of 80th Transportation Research Board Annual Meeting, Washington, DC.*, 2001.

Lowrie, P. The Sydney Coordinated Adaptive Traffic System - Principles, Methodology, Algorithms. IEE Conference on Road Traffic Signaling, *IEE Publication 207*, 1982, pp. 67-70.

Luk, J.Y.K., Sims, A.G. and P.R. Lowrie. SCATS-Application and Filed Comparison with a TRANSYT Optimized Fixed Time System. *Conference of Institute of Electrical Engineers*, Vol. 207, 1982, pp. 71-74.

Martin, Peter T. and Stephen L.M. Hockaday. SCOOT-An Update, *Journal of Institute of Transportation Engineers*, Vol. 65, No. 1, Jan, 1995, pp. 44-48.

Memon G.Q. and A.G.R. Bullen. Multivariate Optimization Strategies for Real time Traffic Control Signals. *Transportation Research Record*, No. 1554, 1996, pp. 36-42.

Muth, Eginard J. (1977). Transform Methods with Application to Engineering and Operations Research. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Ogata, Katsuhiko. (1970). Modern Control Engineering. Englewood Cliffs: New Jersey: Prentice-Hall.

Ogata, Katsuhiko. (1990). Modern Control Engineering. Englewood Cliffs, New Jersey: Prentice-Hall.

Oh, Heung-Un and Sisiopiku, P. Virginia. A Modified ALINEA Ramp Metering Model. *Proceedings of 80th Transportation Research Board Annual Meeting, Washington, DC.*, Jan 2001.

Owen Larry E. and Charlie M. Stallard. A Rule Based Approach to Real-Time Distributed Adaptive Signal Control. *Transportation Research Record*, No. 1683, 1998, pp. 36-42.

Papageorgiou, M., Hadj-Salem, H., and J. Blosseville, ALINEA: A Local Feedback Control Law for On-Ramp Metering. *Transportation Research Record*, No. 1320, 1991, pp. 58-64.

Papageorgiou, M. An Integrated Control Approach for Traffic Corridors. *Transportation Research-C*, Vol. 3, No. 1, 1995, pp. 19-30.

Pavlis, Y. and M. Papageorgiou. Simple Decentralized Feedback Strategies for Route Guidance in Traffic Networks. *Journal of Transportation Science*, Vol. 33, No. 3, 1999, pp. 264-278.

Robertson, D. I. and David R. Bretherton. Optimizing Networks of Traffic Signals in Real Time. *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 1, Part 1, Feb 1991, pp. 11-15.

Shimizu, H., H. Kita. Feedback Control of Signal Control Parameters for a Traffic Network. *Abstracts of the Third World Congress on Intelligent Transport Systems,* 1996.

Shimizu, H., Mashiba, H., and N. Inoue. A Systematic Signal Control of Congestion Lengths. *Proceedings of IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems,* 1999, pp. 164 -169.

Shimizu, H. and H. Mashiba. Signal Control of Two-Way Traffic Arterials. *Proceedings of the 39th SICE Annual Conference,* International Session Papers, 2000, pp. 301-306.

Stanaitis, O.E. (1967). An Introduction to Sequences, Series, and Improper Integrals. San Francisco: Holden-Day, Inc.

Su, C.C. and G.L. Chang. An Integrated Model for Adaptive Intersection Signal Control. *Intelligent Transportation Systems Journal.* Vol. 3, Issue 3, 1996, pp. 225-248.

Tham, M.T. University of New Castle Upon Tyne. Chemical Engineering and Advanced Materials. http://www.psrc.org/datapubs/pubs/trends/t9trend.htm. (5th May 2002).

Wang, Yibing and Markos Papageorgiou. Feedback Routing Strategies for Freeway Networks: A comparative Study. *Journal of Traffic and Transportation Studies,* No. 2, 2000, pp. 642-649.

Wey, W.M., Jayakrishnan, R., and M. G. McNally. A Local Feedback Controller for Oversaturated Intersection Control Based on Dynamic Road Traffic Models. *Proceedings of 6th International Vehicle Navigation and Information Systems Conference in conjunction with the Pacific Rim TransTech Conference,* 1995, pp. 172-178.

Wey, Wann-Ming and R. Jayakrishnan. Network Traffic Signal Optimization Formulation with Embedded Platoon Dispersion Simulation, *Transportation Research Record,* No. 1683, 1999, pp. 150-159.

Wey, Wann-Ming and R. Jayakrishnan. Performance Studies of a Network Adaptive Traffic Control Algorithm via Simulation Model. *IEEE Intelligent Transportation Systems Conference proceedings,* August 25-29, 2001, pp. 663-668.

Wey, Wann-Ming. Model Formulation and Solution Algorithm of Traffic Signal Control in an Urban Network, *Journal of Computers, Environment and Urban Systems,* Vol. 24, 2000, pp. 355-377.

Wolshon, P. and C.T. Willam. Analysis of Intersection Delay under Real-Time Adaptive Signal Control. *Transportation Research Part C,* Vol. 7, No. 1, Feb 1999, pp. 53-72.

Wood, K., Palmer, P., and R.D. Bretherton. Congestion Analysis and Diagnosis in UTC Networks. *IEE Conference Publication*, No. 391, 1994, pp. 172-176.