SOME ASPECTS OF OPTIMISATION OF TRAFFIC SIGNAL TIMINGS FOR TIME-VARYING DEMAND

by

Bin Han

A thesis submitted to the University of London

for the Degree of Doctor of Philosophy

Transport Studies Group

University College London

May 1990

ABSTRACT

The literature of applications of mathematical optimisation to the problem of setting fixed-time traffic signals at individual road junctions is reviewed. The methods developed so far have been confined largely to steady-state conditions. The research described in this thesis therefore aims to develop an optimisation technique to solve the problem of setting traffic signals for periods of time-varying demand.

Previous methods of optimising traffic signal timings with respect to the total delay in the junction in a single time period are unsatisfactory especially when the junction is oversaturated. The existing program OSCADY, although it assesses performance of the junction by means of the sheared queueing formulae, optimises with respect to other criteria. Optimisation of signal timings with respect to the existing sheared delay formulae gives rise to difficulties in optimising a non-convex, non-differentiable objective function. A new sheared delay formula has been developed in order to overcome these difficulties. The properties of this formula have been investigated and optimisation of signal timings with respect to this new formula has been implemented. The results have been compared with those given by the previous method SIGSET and an approximate method that avoids the difficulties in optimisation. The comparison shows that the new formula is appropriate for use in optimising traffic signals for a single period.

When traffic demands are different in successive periods, the signal settings that are optimal for each individual period, as given for example by the program OSCADY, are only local solutions to the problem. These settings may be readjusted and the changes between them shifted in time so that the overall performance for those periods taken together is improved. A sequential optimisation technique has been developed to minimise the total junction delay over the successive periods taken together by searching for the optimal signal timings and the time-shifts subject to certain queue length considerations. Some example calculations are made and the results show that such a technique can provide modest improvements in the junction control performance and give less delay than the existing methods that optimise only for individual periods.

Suggestions for further research are given, together with a list of references in this field.

ACKNOWLEDGEMENTS

I would like to thank the following:

My supervisor Professor Richard Allsop, for his advice, patience, understanding and encouragement throughout the course of the research. The completion of this work would not be possible without his constant supervision and help.

Dr. Benjamin Heydecker, for his invaluable advice and suggestions.

Other colleagues in the Transport Studies Group, past and present, especially Dr Pilo Willumsen, Mr Andrew Sang, Mr Paul Silcock, Ms Gladys Frame, Mr Jamshid Aziminia, Mr Fernando Iunes, Mr Kenneth Wong and Mrs Saskia Fry, for their help and encouragements in many respects.

The Chartered Institute of Transport, for awarding me the Henry Spurrier Travelling Scholarship in 1988, which enabled me to have an opportunity to visit two West German institutions where Professors Leutzbach and Retzko offered a lot of help during the visit.

The British Council and the Chinese Government, for supporting the work financially.

Finally, I would like to thank my wife Lingwei, for her consideration, patience, and devoted sacrifices.

CONTENTS

PAGE

ABSTRACT 2

ACKNOWLEDGEMENT 3

CHAPTER 1. INTRODUCTION

§1.1 BASIC CONCEPTS OF TRAFFIC SIGNAL CONTROL 9
§1.2 DEFINITIONS 10
§1.3 CONSTRAINTS IMPOSED ON SIGNAL TIMINGS 14
§1.4 NOTATION 15
§1.5 EVALUATION OF JUNCTION PERFORMANCE 17
§1.6 TASKS INVOLVED IN DESIGNING FIXED-TIME TRAFFIC SIGNALS 19
§1.7 TWO KINDS OF CONTROL 19
§1.8 ORGANISATION OF THE THESIS 20

CHAPTER 2. EXISTING METHODS OF SETTING TRAFFIC SIGNALS FOR A SINGLE JUNCTION

§2.1 INTRODUCTION 21

§2.2 STEADY-STATE THEORETICAL MODELS 22

§2.2.1 Introduction 22
§2.2.2 Various Models For Estimating The Delay Per Pcu 22
§2.2.3 The General Model For Estimating The Delay Per Unit Time 24

§2.3 DETERMINSTIC THEORY AND THE COORDINATE TRANSFORMATION 25

§2.3.1 Introduction 25
§2.3.2 Determinstic Models 26
§2.3.3 The Development of Comprehensive Formulae 26

§2.4 TRAFFIC SIGNAL SETTINGS FOR A SINGLE JUNCTION 30

§2.4.1 The Undersaturated Junctions 30
§2.4.2 Junctions That May Be Oversaturated 34

§2.5 THE PHASE-BASED CONTROL METHOD 35

§2.5.1 Introduction 35
§2.5.2 Previous Work 36
§2.5.3 The Variables And Constraints 38
§2.5.4 Optimisation Of Signal Timings 41

§2.6 COMPUTER PROGRAMS FOR FIXED TIME TRAFFIC SIGNALS 42

§2.6.1 Introduction 42
§2.6.2 Program OSCADY 43

§2.7 CONCLUSION 45

CHAPTER 3. ESTIMATION OF QUEUE LENGTHS AND DELAYS AT TRAFFIC SIGNALS

§3.1 INTRODUCTION 46
§3.2 QUEUE LENGTH ESTIMATION 47

§3.2.1 Random Queue Length Estimation 47
§3.2.2 Uniform Queue Length Estimation 54

§3.3 DELAY ESTIMATION 60

§3.3.1 Random Delay Estimation 60
§3.3.2 Uniform Delay Estimation 63

§3.4 CONCLUSION 67

CHAPTER 4. TRAFFIC SIGNAL SETTINGS FOR A SINGLE PERIOD

§4.1 INTRODUCTION 69
§4.2 DIFFICULTIES IN OPTIMISATION 70
§4.3 AVOIDING THE DIFFICULTIES USING EXISTING DELAY EXPRESSIONS 71

§4.4 A NEW APPROACH 73

§4.4.1 Introduction 73
§4.4.2 New Expression For Delay 73
§4.4.3 Acceptability And Convexity 81

§4.5 THE OPTIMISATION METHOD 83
§4.6 EXAMPLES 85

§4.6.1 Introduction 85
§4.6.2 Junction 1 — Symmetrical Crossroads of Two One-way Streets 86
§4.6.3 Junction 2 — Asymmetrical Crossroads of Two One-way Streets 89
§4.6.4 Junction 3 — Chapel Hill Junction 95

§4.7 CONCLUSION 101

CHAPTER 5. TRAFFIC SIGNAL SETTINGS FOR PERIODS OF TIME-VARYING DEMAND

§5.1 INTRODUCTION 103
§5.2 STATEMENT OF THE PROBLEM 104
§5.3 NOTATION 106
§5.4 MATHEMATICAL FORMULATION 110

§5.4.1 Basic Assumptions On Signal Settings 110
§5.4.2 The Calculation Of The Performance Index 111
§5.4.3 Evaluation Of The Effects Of $(\underline{\lambda_k}')$ And (δT_k) On The Performance Index 117

§5.5 OPTIMISATION OF $(\underline{\lambda_k}')$ FOR FIXED (δT_k) 118

§5.5.1 Introduction 118
§5.5.2 The Calculation Of The Derivatives With Respect To $\underline{\lambda_k}'$ 118
§5.5.3 The Sequential Optimisation Of $(\underline{\lambda_k}')$ 120
§5.5.4 An Example Calculation 121

§5.6 OPTIMISATION OF (δT_k) FOR FIXED ($\underline{\lambda_k}'$) 126

§5.6.1 Introduction 126
§5.6.2 The Example Calculations 126
§5.6.3 The Optimisation Method - Golden Section Method 130
§5.6.4 Results For The Present Example 132

§5.7 A SEQUENTIAL OPTIMISATION METHOD 134

§5.7.1 Introduction 134
§5.7.2 The Algorithm 134
§5.7.3 Results For The Present Example 135

§5.8 TAKING INTO ACCOUNT THE SUBSEQUENT CONDITIONS 137

§5.8.1 Introduction 137
§5.8.2 A Further Algorithm 138
§5.8.3 Results For The Present Example 139

§5.9 MORE EXAMPLE CALCULATIONS FOR TWO-PERIOD CASES 142

§5.9.1 Introduction 142
§5.9.2 Junction 1 — Symmetrical Crossroads 142
§5.9.3 Junction 2 — Asymmetrical Crossroads 146
§5.9.4 Junction 3 — Chapel Hill Junction 152
§5.9.5 Discussions 156

CHAPTER 6. CONCLUDING REMARKS AND SUGGESTIONS FOR FURTHER WORK

§6.1 CONCLUDING REMARKS 157
§6.2 SUGGESTIONS FOR FUTURE WORK 158

§6.2.1 Further Investigations of Time-shift 158
§6.2.2 Multi-criteria Optimisation 158
§6.2.3 Phase-based Optimisation 159
§6.2.4 Wider Experience 159

REFERENCES 160

APPENDIX 1: THE SOLUTION OF A CUBIC EQUATION 165

APPENDIX 2: THE DERIVATIVES OF THE NEW DELAY FORMULA 169

APPENDIX 3: THE DERIVATIVES OF THE RANDOM QUEUE LENGTH FORMULA 173

CHAPTER 1. INTRODUCTION

§1.1 BASIC CONCEPTS OF TRAFFIC SIGNAL CONTROL

The ever increasing car ownership throughout the world has significantly improved the quality of our lives. However, as a mixed blessing, it has also caused many serious problems such as traffic congestion, environmental pollution, and road accidents. To deal with these difficulties, a lot of efforts have been made, among which is the installation of numerous traffic signals. They are usually used to allocate the right of way to the conflicting traffic streams at busy urban street junctions, so that the traffic can cross safely and the road system can be used more effectively, and by proper choice of the signal settings, the delay to the traffic, the number of stops and hence the fuel consumption can be reduced. They can also help to provide crossing facilities for other road users like pedestrians and cyclists, and to give priority to public transport and emergency service vehicles.

There are two main types of traffic signals: fixed-time and vehicle actuated. With <u>fixed-time signals</u> the settings are predetermined and of fixed duration within a certain period of time, whereas the operation of <u>vehicle-actuated signals</u> is controlled partly or fully by traffic approaching the junction, but in this kind of control, a preset schedule is also required when there is heavy traffic on all roads approaching the junction, especially if the junction is overloaded, or when faults occur in the vehicle actuation apparatus. In the cases where the neighbouring different junctions are close enough, the signals at junctions can be linked so that their operation is co-ordinated. Since the controllers for fixed-time signals are relatively simple and cheaper, there are still many of them in use. Research on fixed-time traffic signals is still very important nowadays, both for the operation of such signals and for the preset schedules required by vehicle actuated signals.

In this thesis, the focus is put on fixed-time traffic signals at an isolated junction, i.e. a junction whose signal operation is not linked with the signals at other junctions.

§1.2 DEFINITIONS

It is assumed in this thesis that the sequence of signal aspects is red, red-and-amber, green, and amber. This is the common practice in Britain.

The roads meeting at a signalised junction are usually divided into approaches. An approach can be regarded as an area of road leading to the junction and such that traffic waiting there to cross forms only a single queue (or behaves like a single queue). Therefore a road entering an junction may have one or more approaches depending on whether all the traffic is following the same direction or there are different queues for each direction, and an approach may consist of several lanes. An approach has right of way when traffic from it is allowed to pass the signal. Similarly, a stream of traffic comprises either one lane or several adjacent lanes of traffic which behaves as a single queue independent of behaviour of traffic in any other adjacent lanes. A stream is the smallest subdivision of traffic that need to be considered in calculating signal timings. Two or more traffic streams are called compatible if they can receive green indications simultaneously and therefore they may safely cross the junction simultaneously. The arrival rate of traffic in each stream is used to specify the pattern of traffic for which timings are to be calculated.

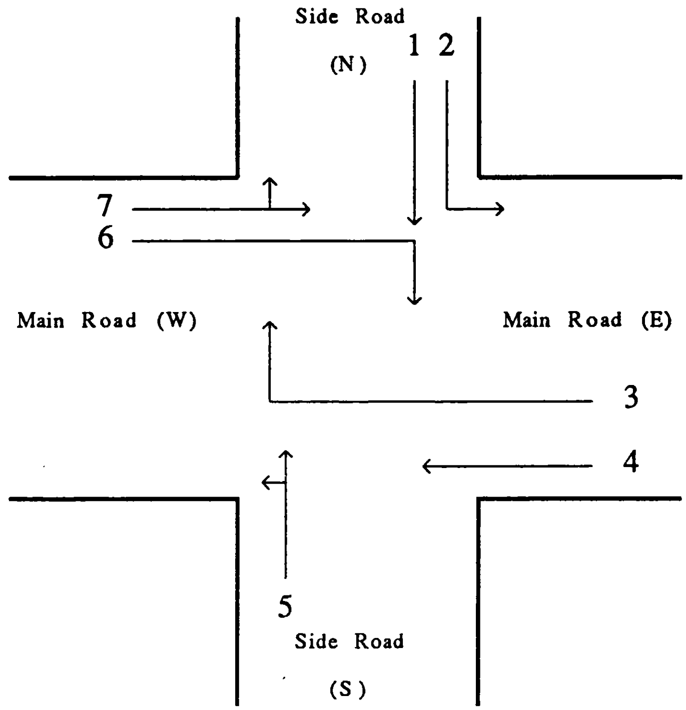
In fixed-time operation of traffic signals, two or more sets of mutually compatible streams receive green in turn for specified times. One repetition of this process is called a signal cycle and its duration the cycle time. In traffic engineering practice, for each stream a signal cycle is usually divided into alternate periods called effective red and effective green times for that stream. It is assumed that in the effective red times, traffic in the stream cannot pass the signal, and in the effective green periods, it passes the signal at a uniform rate called the saturation flow so long as there is a queue in the stream and passes as it arrives if there is no queue. The effective green period is related to but not the same as the green period that is actually displayed by the signals. The relationship depends on the behaviour of traffic in the stream and the junction layout, and may therefore differ among streams receiving the same signal indications. The value for saturation flow may be measured or estimated by using the relevant physical layout information for the stream (eg. width of the corresponding approach, number of lanes, gradient, etc), and details are given in TRRL report RR67 (**Kimber** et al

1986).

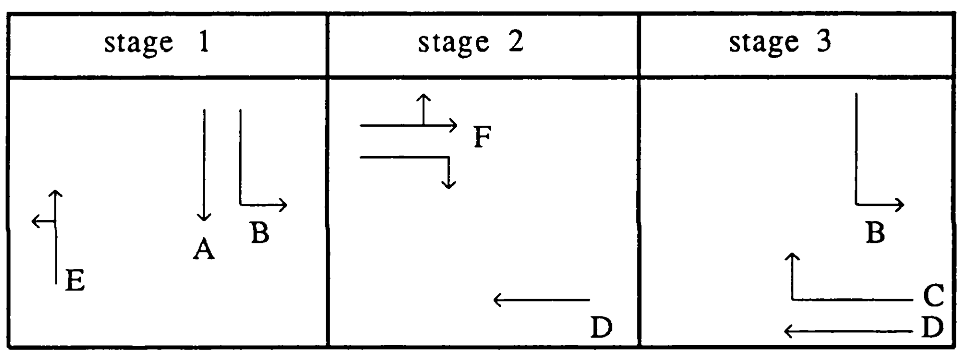
A set of streams which always receive identical signal displays is called a group. Two groups are said to be compatible if every stream in one group can safely have green at the same time as every stream in the other. If two groups are incompatible, a clearance time must elapse between the end of displayed green for one and the start of displayed green for the other. A phase is the sequence of signal conditions given to one or more vehicle or pedestrian streams of traffic so that each stream allocated to the phase receives identical signal indications. A part of the cycle in which a particular set of groups has green is called a stage. The arrangement for the change between stages subject to the clearance time requirements is the interstage structure. Usually there is more than one practicable interstage structure. The composition of the stages and their order in the cycle together form a sequence.

For any given sequence and interstage structures, the duration of each stage can be defined as the intersection of the displayed greens of the groups having green in the stage. The time between one stage and next is called the interstage time which is determined by the clearance times and the interstage structure. The time from the end of the effective green time for one stage to the beginning of that for the next is called the lost time following the former stage. The sum of the lost times over all the stages is called the total lost time per cycle. The proportion of the cycle time that is effectively green for a stage is called the effective green time for that stage. Some of the above definitions can be illustrated by Fig 1.1, where a typical cross-roads is described.

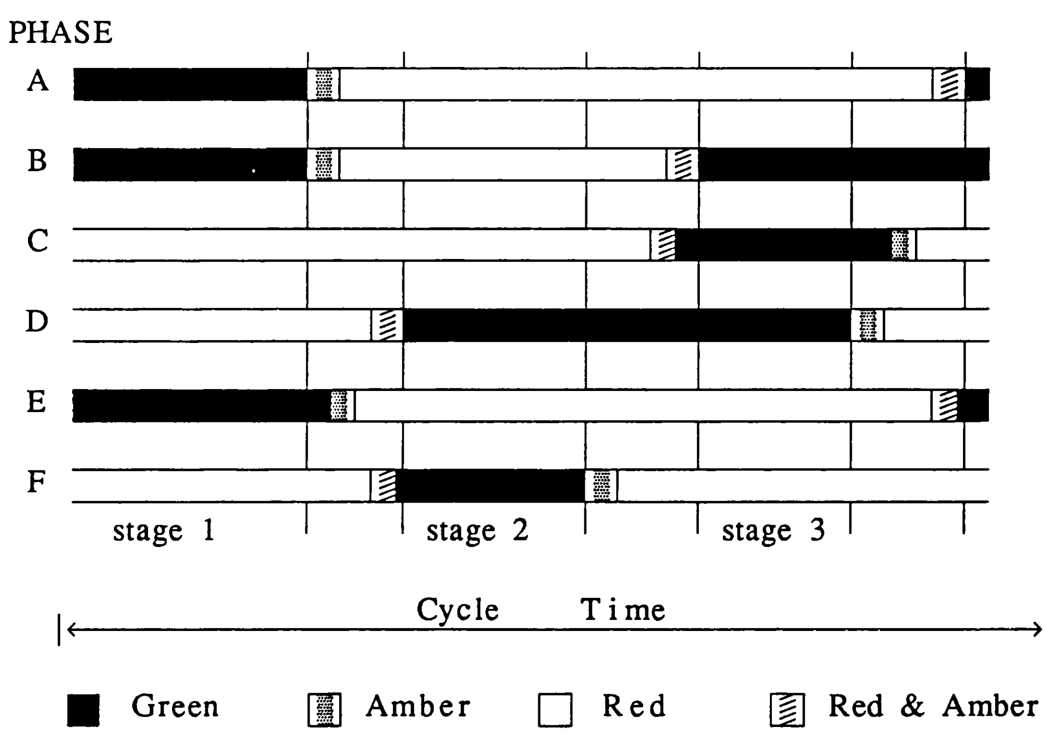
However, to establish the relationship between the effective green time for each stream and the effective green times for those stages in which this stream has right of way, the concept of extra effective green time for a stream needs to be introduced. If a stream does not have right of way in every stage, its extra effective green time is the sum of the amounts by which the effective green time for that stream overlaps the lost times before the first and after the last stage in which the stream has right of way. For a stream that has right of way in every stage then its extra effective green time is minus the amount by which the cycle time exceeds the effective green time for that stream.



Side Road
(N)
1 2
7
6
Main Road (W)
Main Road (E)
3
4
5
Side Road
(S)



stage 1
stage 2
stage 3
E
A
B
F
D
B
C
D



PHASE
A
B
C
D
E
F
stage 1
stage 2
stage 3
Cycle Time
Green
Amber
Red
Red & Amber

Fig 1.1

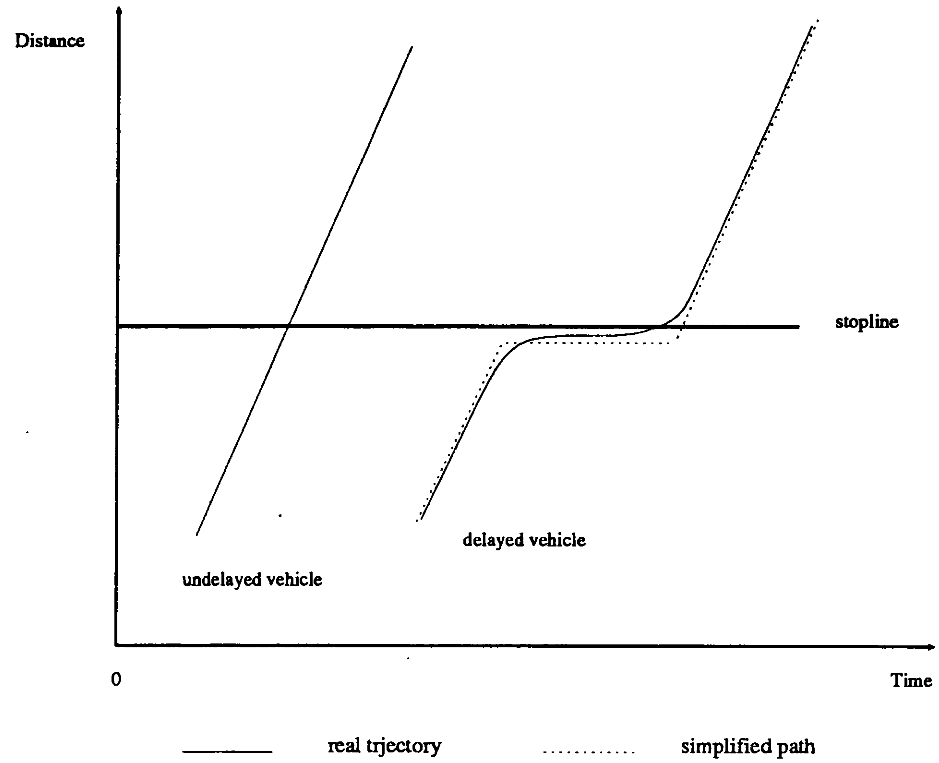


Fig 1.2 The Delay To A Vehicle

There are several measures of the performance of the signal timings for a stream. The capacity of the stream is the maximum long term rate at which traffic can enter the junction. The difference between the travel time of a vehicle and the time it would have taken if the prevailing level of traffic were present and the stream in which it travels had continuous right of way is called the delay incurred by this vehicle.

In order to make calculations of delay, the delay can be considered as occurring at the stopline, with the delayed vehicles travelling at undelayed speed until they reach the stopline where they decelerate instantaneously, spending their period of delay there before accelerating instantaneously back to undelayed speed to pass the junction. This can be illustrated by Fig 1.2.

Based on this assumption, the delay can further be regarded as coming from three sources:

(a) *Randomness* in the arrival of traffic.

(b) *Overload* when the degree of saturation is greater than one for some time.

(c) *Alternation of red and green* which causes delay even to traffic that arrives uniformly.

Delay arising from (a) and (b) is called random delay, and addition delay caused by (c) is called uniform delay. They can be estimated separately. To measure the delays in a stream, two measurements can be used: the first is the average delay per pcu (or per vehicle), d, and the second is called the delay per unit time, D, which is regarded as the average of the number of vehicles that would be queueing at any time instant during a certain time period if the vehicles followed that simplified paths. This is preferable for use when time-dependent expressions for delays are employed.

Correspondingly, the traffic queues occurring at a signalised junction can be regarded as consisting of two parts: the random queue length and the uniform queue length, and they are also estimated separately. Formulae for estimating queues and delays will be given explicitly in Chapter 3.

When estimating delay or queue lengths, the traffic is usually regarded as consisting of identical vehicles equivalent in size and performance to an average passenger car. Traffic flows are thus expressed in terms of these passenger car units (known as pcu's) and actual vehicles of various types are considered as equivalent to different numbers of these units.

§1.3 CONSTRAINTS IMPOSED ON SIGNAL TIMINGS

To ensure safe and practical operation of fixed-time signals, some constraints have to be introduced. These usually include the following:

1) *Clearance time constraints:*

A clearance time must elapse between any consecutive green times for incompatible groups.

2) *Cycle time constraint:*

The cycle time should not be too long or too short, therefore a maximum or minimum cycle time or both may be introduced, or the cycle time can be specified.

3) *Green time constraints:*

Maxima or minima or both may be imposed on the green times for various phases or stages or for both.

4) *Maximum acceptable degree of saturation:*

There should be a maximum acceptable value of X, the degree of saturation, for each stream. This can be denoted by P, and the product of P and the capacity of a stream is then called the practical capacity for that stream.

5) *Queue length constraints:*

Due to the physical layout of the junction, the maximum queue length at the end of the red for certain streams may be subject to an upper limit, or some penalty may be imposed on unnecessarily longer queues.

§1.4 NOTATION

The following commonly accepted notation will be used throughout this thesis. Let

g = the effective green time for a stream in seconds
r = the effective red time for a stream in seconds
e = the extra effective green time for a stream in seconds
c = the cycle time in seconds ($c = r + g$)
λ = proportion of the cycle that is effectively green for a stage
Λ = the proportion of the cycle that is effectively green for a stream ($\Lambda = g/c$)
C = a parameter depending on the service patterns ($C = 0.5 \sim 0.6$)
q = the average arrival rate for a stream in pcu/second
s = the saturation flow rate on a stream in pcu/second
y = the flow ratio of this stream ($= q/s$)
Q = the traffic capacity of a stream ($= \Lambda s$)
X = the degree of saturation for a stream ($= q/Q$)
P = the maximum acceptable X for a stream
L = the queue length in pcu or the lost time per cycle in seconds (whichever applicable)
t = time instant
T = the length of a time period over which q and Q are assumed constant
d = the average delay to a pcu in a stream in seconds

D = the delay per unit time (rate of delay) in pcu

W= the total delay over the period T (=DT)

μ = the common factor by which q can be increased in each stream

σ = the reserve capacity as a percentage of the given arrival rate

Some subscripts will be used so that different variables can be represented by combining them with the above basic letters. The following rules apply unless specified. If there are n streams in the junction, and there are m stages in the signal cycle and there are p time periods during each of which the traffic flow in each stream is constant, then let

i denotes in stage i ($i = 1, 2, \cdots, m$)

j denotes in stream j ($j = 1, 2, \cdots, n$)

k denotes during or at the end of (whichever applicable) the kth time sub-period ($k = 1, 2, \cdots, p$)

r denotes the random component

u denotes the uniform component

e denotes equilibrium

o denotes oversaturation

0 denotes initial

max denotes maximum value

min denotes minimum value

And the stage matrix can be defined as

$$A = (a_{ij}) \qquad (j = 1, 2, \cdots, n;\ i = 1, 2, \cdots, m)$$

Where

$$a_{ij} = \begin{cases} 1 & \text{if stream } j \text{ has green in stage } i \\ 0 & \text{if not} \end{cases}$$

Also define

a_{0j} = proportion of the lost time that is effectively green for stream j. ($j = 1, 2, \cdots, n$)

λ_0 = total lost time over the cycle time, i.e. $\lambda_0 = L/c$

In the stage-based method (definition given below in §1.7), the proportion of the effective green time for stream j in the junction can then be expressed as:

$$\Lambda_j = \sum_{i=0}^{m} a_{ij}\lambda_i \qquad (j = 1, 2, \cdots, n;\ i = 1, 2, \cdots, m) \qquad (1.1)$$

§1.5 EVALUATION OF JUNCTION PERFORMANCE

To evaluate the performance of a signalised junction, a number of different criteria can be applied. One criterion is known as the reserve capacity denoted by σ, and can be given by:

$$\sigma = 100\cdot(\mu-1)\% \qquad (1.1a)$$

Where μ is a common factor which can be defined as the maximum positive multiplier that can be applied to the arrival rate in each stream whilst satisfactory operation can be assured. If $\mu > 1$ (i.e. $\sigma>0$), then the junction is said to be within capacity; If $\mu \leq 1$(i.e. $\sigma\leq0$), then the junction is overloaded. For known signal settings, σ can also be given by:

$$\sigma = 100\cdot\left\{\mathrm{Min}\left(\frac{P_j}{X_j}\right)-1\right\}\% \qquad (j=1,2,\cdots,n)$$

$$=\mathrm{Min}(\sigma_j) \qquad (1.1b)$$

where $\sigma_j = 100\cdot\left\{\left(\frac{P_j}{X_j}\right)-1\right\}\% \qquad (j=1,2,\cdots,n) \qquad (1.2)$

and σ_j is called the reserve capacity for stream j.

In most cases, however, the total rate of delay is the most commonly used criterion. This is the sum over all streams of the delay per unit time, i.e.

$$D = \sum_{j=1}^{n} D_j \qquad (1.3)$$

Or alternatively, the total delay over a certain period of time for a stream can be defined as:

$$W_j = D_j T \tag{1.4}$$

Hence the total delay over T for the junction can be defined as:

$$W = \sum_{j=1}^{n} W_j \tag{1.5a}$$

or

$$W = DT \tag{1.5b}$$

Where D is defined by (1.3).

Formula (1.5b) shows that to set traffic signals for a single time period T during which all the q are constant, whether to use the total rate of delay D or to use the total delay over the period, W, as a performance index, will make no difference in the resulting signal settings.

In addition, other criteria may include:

(a) The degree of saturation, which is defined as the ratio of flow rate to capacity for a particular stream, i.e. $X = q/Q$. If $X < 1$, then the stream is undersaturated; whereas if $X \geq 1$, then the stream is oversaturated. Therefore for a junction, if all the streams are undersaturated, then the junction is undersaturated; However, if at least one of the streams is oversaturated, the the junction is oversaturated.

It should be pointed out that the terms *oversaturated* and *overloaded* are related but not necessarily the same. If the maximum acceptable degree of saturation $P = 1$, then they are the same, i.e. an oversaturated junction can also be said to be an overloaded junction; But if $P < 1$ (e.g. $P=0.9$), then they are different: a junction may for certain arrival rates be overloaded but also be undersaturated.

(b) The critical cycle time, which is the shortest cycle time that will ensure each stream operates within its practical capacity;

(c) Stop rate, i.e. average number of vehicles that have to stop per unit time;

(d) Fuel consumption;

(e) Queue lengths.

§1.6 TASKS INVOLVED IN CALCULATING SIGNAL SETTINGS

Generally, the following factors should be considered when calculating signal settings either for a given signalised road junction or in the case of designing such a junction:

(a) Determination of the clearance times required for the conflicting groups;
(b) Choice of stage composition;
(c) Choice of stage sequence and interstages;
(d) Calculation of cycle time and green times with respect to one or more of the performance criteria in §1.5;
(e) Performance evaluation of the junction as a consequence of the timings calculated in (d).

§1.7 TWO KINDS OF CONTROL

Two kinds of control have been used at single road junctions. In stage-based control, the duration of stages are considered as the main variables, and only (d) and (e) are involved in the calculation process, whereas (a), (b) and (c) are supposed to be predetermined. Hence it has some disadvantages, e.g. it requires the user to specify the stage sequence, the stage composition and the duration and details of the transition periods. In the more flexible approach called phase-based control, however, all the above factors ((a)-(e)) are considered and calculated together, so that the control efficiency can be enhanced.

The new methods described in this thesis are equally relevant to both approaches, but have been developed and tested using the stage-based approach to signal optimisation because methods based on it are better established and easier to implement and adapt. The literature review given in the next Chapter also extends to the phase-based approach, which has been fully implemented only concurrently with the work described here.

§1.8 ORGANISATION OF THE THESIS

Chapter 1 gives a brief introduction to the fundamentals of traffic signals and the basic definitions and notation.

Chapter 2 reviews the existing methods of setting traffic signals for a single junction, most of which are stage-based, but a survey is also made for the more advanced phase-based method. The literature review suggests that to set traffic signals for a single time period with respect to minimum total delay, these methods are unsatisfactory especially when the junctions may be oversaturated. When traffic demands are different in successive periods, only the program OSCADY can model the situation but even this can only optimise for each periods individually.

Chapter 3 studies the various formulae available for estimating queues and delays. Those formulae that are appropriate in different situations are identified. The results suggest that the formulae for estimating queues and delays used in OSCADY have some deficiencies, although they will be used for evaluation in the subsequent chapters for the convenience of comparing the outputs of the present method with OSCADY.

Chapter 4 discusses the problem of setting traffic signals for a single time period. The limitations of the existing methods and the difficulties of further improvement are identified. A new delay expression is established to overcome these difficulties. Some example calculations show that this approach is appropriate for optimising traffic signals for a single period.

Chapter 5 focuses on the problem where there is more than one time period to be modelled during each of which the traffic flow for each stream is regarded as constant. A sequential optimisation technique taking into account the subsequent delays after the end of the whole time periods is developed which will in most cases finds the global optimum of the signal settings. A few example illustrations suggest that such a method can give somewhat better results than the previous methods like OSCADY.

Chapter 6 gives some concluding remarks and some suggestions for further work, followed by a list of *References*.

CHAPTER 2 EXISTING METHODS OF SETTING TRAFFIC SIGNALS FOR A SINGLE JUNCTION

§2.1 INTRODUCTION

The traditional methods of calculating the optimal signal settings for an isolated junction are mostly stage-based. As mentioned above, this kind of control needs the predetermination of the stage composition and the stage sequence according to the junction layout and the requirements for safety and convenience. The optimisation process only involves the calculation of cycle time and the allocation of green times to each stage, therefore to each traffic stream.

There are different criteria that can be used as a performance index in the calculation of optimal signal timings, e.g. average total rate of delay, total delay, the reserve capacity of the junction, the stop rate, or fuel consumption. Very often delay minimisation is performed, but sometimes other criteria are also needed, and combinations of them may be used in some cases.

According to the assumptions made and the range of application, the theoretical background for the various approaches could until the last ten years or so be divided into two categories: the steady-state stochastic theory and the deterministic theory. The former is conventional and can be used to solve the problems of setting traffic signals for undersaturated junctions where an equilibrium condition usually exists. The deterministic queueing theory, on the other hand, can deal with the situations when the junctions are substantially overloaded. But when the junction operates nearly at or just over their capacity, which is the most important region of operation during the peak period or even for normal traffic condition, both of the above theories fail to give satisfactory results. A relatively recent approach, based on an idea proposed by **P.D. Whiting**, has been developed by various authors to establish a so-called sheared delay formula to treat the whole range of traffic demand for estimating the random delay, including variation over time. But so far there has been no single comprehensive formula available for estimating the total delay (random + uniform) for the whole range of the degree of saturation.

The literature on the fixed-time control of a single junction with the more advanced phase-based approach is also reviewed in this thesis, although the framework for this thesis is set up under the stage-based approach.

§2.2 STEADY-STATE THEORETICAL MODELS FOR ESTIMATING DELAYS

§2.2.1 Introduction

For an undersaturated junction, the traffic will reach a steady state or equilibrium. Theoretical studies of fixed-time control for a single junction that is undersaturated have been made by a lot of researchers (e.g. **Clayton** 1941, **Wardrop** 1952, **Webster** 1958, **Miller** 1963, 1968, **Webster** and **Cobbe** 1966, **Allsop** 1970, 1971, 1972, 1976, **Ohno** 1978, etc). Different expressions for vehicular arrival model, departure model and delay have been derived. Mainly there are three kinds of arrival models: Regular Arrivals, Binomial Arrivals and Poisson Arrivals. A detailed analysis of these expressions has been given by **Allsop** (1972) and numerical comparisons has also been made by **Hutchinson** (1972). However, it is a common practice to assume that the arrival process is Poisson, and the departure of the vehicles happens at equal time intervals $\delta t=1/s$ (s is the saturation flow rate) so long as there is still a queue, and the departure time of the first vehicle coincides with the start of the effective green time.

§2.2.2 Various Models For Estimating The Delay Per Pcu

In the past a number of formulae have been derived for the calculation of delay, capacity and queue length, and the theoretical results have been in many cases compared with field data. Among the different expressions for estimating delay at a signalised junction, the most important ones are those of **Webster** (1958), **Miller** (1963), **Newell**(1965) and **Ohno** (1978). **Webster's** full expression based on a Poisson arrival model for the average delay per pcu in seconds is:

$$d_e=\frac{c(1-\Lambda)^2}{2(1-\Lambda X)}+\frac{X^2}{2q(1-X)}-0.65\left\{\frac{c}{q^2}\right\}^{1/3}X^{(2+5\Lambda)} \qquad (2.1)$$

Here the subscript e stands for 'equilibrium'.

By making a digital simulation and regression analysis, **Webster** pointed out that the third term in the above expression represented between 5% and 15% of the total d, hence a simplified formula (known as **Webster's** two-term delay formula) can be given by:

$$d_e = \frac{9}{10}\left\{\frac{c(1-\Lambda)^2}{2(1-\Lambda X)} + \frac{X^2}{2q(1-X)}\right\} \tag{2.2}$$

Miller (1963) also used the Poisson arrival model to obtain the following expression:

$$d_e = \begin{cases} \dfrac{(1-\Lambda)}{2(1-\Lambda X)}\left\{c(1-\Lambda) + \dfrac{(2X-1)I}{q(1-X)} + \dfrac{I+\Lambda X-1}{s}\right\} & X \geq 0.5 \\[2ex] \dfrac{(1-\Lambda)}{2(1-\Lambda X)}\left\{c(1-\Lambda) + \dfrac{I+\Lambda X-1}{s}\right\} & X < 0.5 \end{cases} \tag{2.3}$$

where $I = \dfrac{\text{variance of the number of arrivals per cycle}}{\text{mean number of arrivals per cycle}}$

Newell (1965) considered a more general arrival and departure process, derived the formula which also contains the ratio I:

$$d_e = \frac{c(1-\Lambda)^2}{2(1-\Lambda X)} + \frac{IH(\xi)X}{2q(1-X)} + \frac{I(1-\Lambda)}{2s(1-\Lambda X)^2} \tag{2.4}$$

where $\xi = (sg-qc)/(Isg)^{1/2}$ is a dimensionless measure of the spare capacity of the stream, and $H(\xi)$ is a function, obtained by numerical integration, ranging from 1 to zero as ξ varies from zero to infinity.

Ohno (1978) made a stochastic analysis to adapt the formulae by earlier researchers such as **Miller** to obtain several delay formulae, e.g. the modified **Miller's** expression is:

$$d_e = \frac{1-\Lambda}{2(1-y)}\left\{c(1-\Lambda) + \frac{2}{q}N_m + \frac{1}{s} + \frac{1}{s(1-y)}\right\} \tag{2.5}$$

where

N_m =the average overflow queue, which is the average queue at the beginning of the red periods.

It has been shown (**Hutchinson** 1972) that **Webster**'s and **Miller**'s formulae agree quite closely, but **Webster**'s expression has the advantage of algebraic simplicity over the other expressions, and it has been shown to be quite consistent with observed data, it requires only the measurement of arrival rate and saturation flow rate and hence avoids the difficulty of measuring the I ratio. Since a model should be kept as simple as possible, but should also provide a adequate level of accuracy for the estimation of delay, **Webster**'s formula has been widely used both in theoretical analysis and practical work. **Webster**'s two-term delay formula is also convex, which has been proved by **Allsop** (1971) and **Murchland** (1977), both of whom use the positive matrix method, and later by **Gallivan** (1982) who uses the geometric approach.

§2.2.3 The General Model For Estimating The Delay Per Unit Time

As stated above, of all the above mentioned models, the expression (2.1) by **Webster** is the most popular one, but it is established in terms of the delay per pcu, d_e. However, it can also be expressed in terms of delay per unit time, D_c. Corresponding to (2.1), D_c can be expressed as:

$$D_e = qd_e = \frac{9}{10}(D_{re} + D_{ue}) \qquad (2.6)$$

where

$$D_{rc} = \frac{X^2}{2(1-X)} \qquad (2.7a)$$

$$D_{ue} = \frac{qc(1-\Lambda)^2}{2(1-y)} \qquad (2.7b)$$

However, it has since become usual practice to estimate the delay as the sum of D_{re} and D_{ue}, i.e.

$$D_c = D_{re} + D_{ue} \qquad (2.8a)$$

Generally, as pointed out by **Kimber & Hollis** (1979), the random delay per unit time, D_{re}, can be expressed as:

$$D_{re} = \frac{CX^2}{1 - X} \tag{2.8b}$$

Where D_{re} is derived from the classical queueing theory, and D_{ue} is the same as in (2.7).

In (2.8b) C is a parameter depending on the departure pattern; for traffic signals usually C = 0.5 ~ 0.6. Clearly (2.7a) is a special case of (2.8) when C=0.5.

In this thesis, C is taken as 0.6 (for the reason given in §3.4 in Chapter 3).

§2.3 DETERMINISTIC THEORY AND THE COORDINATE TRANSFORMATION

§2.3.1 Introduction

The basic assumption for the steady-state theory is that the junction is operating within capacity and an equilibrium always exists, so that there are no queues accumulating continuously. However, congestion often happens in real life, particularly during peak periods, and in this case the demand will often exceed capacity. Hence formulae such as **Webster's** cannot be used to treat the oversaturation problems, since as the degree of saturation approaches one they predict infinite delay, which is not realistic, and if it is greater than one no steady-state result exists. Although congestion is a complicated problem associated with not only the operation of traffic signals but also many other factors such as the junction layout, proper signalisation may still reduce the oversaturation and even to a minimum: that is the optimal design problem for oversaturated junctions. Deterministic queueing theory can be used to treat this kind of problem, but will seriously underestimate delay when demand and capacity are nearly equal.

The steady-state models involve the estimation of both the random delay and the uniform delay. Since the random delay is the only source for the delay at non-signalised junctions, and is the dominating component of the delays in streams with high degrees of saturation at signalised junctions, in relation to the estimation of random delay a lot of progress has been achieved in the past in both theoretical analysis and practical operation. For example, the coordinate transformation technique makes it possible to establish an approximate model which can be used to estimate the random delay to solve the problems ranging from undersaturation to oversaturation. So far there has been no similar model for estimating the uniform delay or the sum of random and uniform delay.

§2.3.2 Determinstic Models

These models estimate the queues and delays by assuming that the traffic in a stream will arrive uniformly at a constant rate q and departs at a rate Q (Q is the capacity of the stream). For example, the random delay per unit time in a stream can be estimated by the time-dependent expression:

$$D_{ro} = L_{r0} + 0.5(q - Q)t \tag{2.9}$$

And the uniform delay is (see §3.3.2):

$$D_{uo} = 0.5Qc(1-\Lambda) \tag{2.10}$$

Here the subscript o stands for 'oversaturation', since the determinstic expressions are used for oversaturated streams.

Expressions such as (2.9) are quite useful for estimating the random delay when $X > 1$, but they are not applicable to undersaturated situations, due to the limitations of their assumptions and that of their models. For example, the assumption that the random queues grow at a rate determined only by the excess of demand over capacity, i.e. (q-Q), and decay when the demand has fallen below capacity at a rate given by the difference of Q and q ignores the important effects of random fluctuations in arrivals and departures, hence may underestimate delay.

§2.3.3 The Development of Comprehensive Formulae

The limitations of both the steady-state models and the determinstic models raise the need for more powerful models that can be used to treat the full range of the degree of saturation. As a first step, a lot of efforts have been devoted to establishing comprehensive models for estimating the random delay and random queue length, which will be outlined below.

Mayne (1976, 1979) used some results on the transient queueing theory (e.g. **De Smit** 1971) and transformed the existing formulae into formulae using binomial distribution and Poisson distribution, then applied statistical distribution approximations and other techniques of numerical analysis to give comprehensive queue and delay formulae suitable for all degrees of saturation. In this model, a non-zero initial queue length was considered.

Kimber et al (1977) assessed the delays occurring at major/minor priority junctions during peak periods by using the queue length and delay functions developed by applying the coordinate transformation on the basis of zero initial queue lengths. The random arrival/random service model (M/M/1) is employed for the calculations. The transformation was originally suggested by **P. D. Whiting** and is used in calculating delays at traffic signals in TRANSYT 6.

Catling (1977) adapted equations of classical queueing theory (e.g. POLLACZEK-KHINTCHINE equation) to oversaturated traffic conditions and gave time-dependent expressions for delay and queue length which cover the case when there are non-zero initial queues at the start of the time period considered. In **Catling**'s model a parameter C is introduced which is a constant depending on the service time patterns. **Branston** (1978) compared the estimated queue lengths given by **Mayne** and **Catling** with those actually observed during three peak periods in London for different divisions of the peak period. He analysed the sensitivity of estimated queue lengths and delays at oversaturated signal junctions to the methods of making the division. He concluded that both formulae can give good representations of observed data, but they are dependent on the saturation flow. The peak period can normally be divided into a series of about 10 – 15 minute time intervals and C = 0.55 can be used as a most appropriate choice for traffic signals.

Kimber and **Hollis** (1978, 1979) extended the previous results given by different authors and generalised them into formulae suitable for a number of more common cases. They gave detailed derivations of queue length and delay formulae, and the resulting formulae can be applied to treat not only signalised junctions but also roundabouts, priority junctions and motorway merges. Two approaches are used, i.e. high-definition representation which has applications in engineering design and low-definition representation which is applicable to the economic assessment of junction performance.

Akcelik (1982) presented new approximate expressions for predicting delay, stop rate and queue length at signalised junctions in terms of the overflow queue concept. Each expression consists of two parts: the uniform component and the overflow component. His method differs from the previous ones in that he transforms **Miller's** steady-state equation instead of **Webster's** into a time-dependent expression. **Cronje** (1983) treated the traffic flow through a fixed-time signal as a Markov process and developed equations for estimating delay, queue length and stops. Later he (**Cronje** 1986) compared the formulae given by himself and **Mayne** and **Catling** by simulation method, and concluded that his formula is better than the other two.

Kimber and **Hollis's** expressions for random delay and random queue length, currently the most widely-used in British practice, are based on the so-called coordinate transformation technique. For example, to obtain the comprehensive formula for estimating the random delay per unit time, first consider the random delay expressions by the steady-state theory and the determinstic theory:

$$D_r = \begin{cases} D_{re} = \dfrac{CX_c^2}{1-X_c} & X_c<1 \quad \text{(Steady-state theory)} \quad (2.11a) \\ D_{ro} = L_{r0} + 0.5Qt(X_o-1) & X_o \geq 1 \quad \text{(Determinstic theory)} \quad (2.11b) \end{cases}$$

From (2.11b) we have:

$$X_o = \frac{2D_{ro} - 2L_{r0}}{Qt} + 1 \qquad (2.12)$$

In Fig. 2.1, let the line segments AA′ and BB′ be equal, i.e. let

$$1 - X_e = X_o - X \qquad (2.13)$$

Where X is the degree of saturation in the expression for D_r.

Hence

$$X_e = X + 1 - X_o = \frac{qt + 2L_{r0} - 2D_{ro}}{Qt} \qquad (2.14)$$

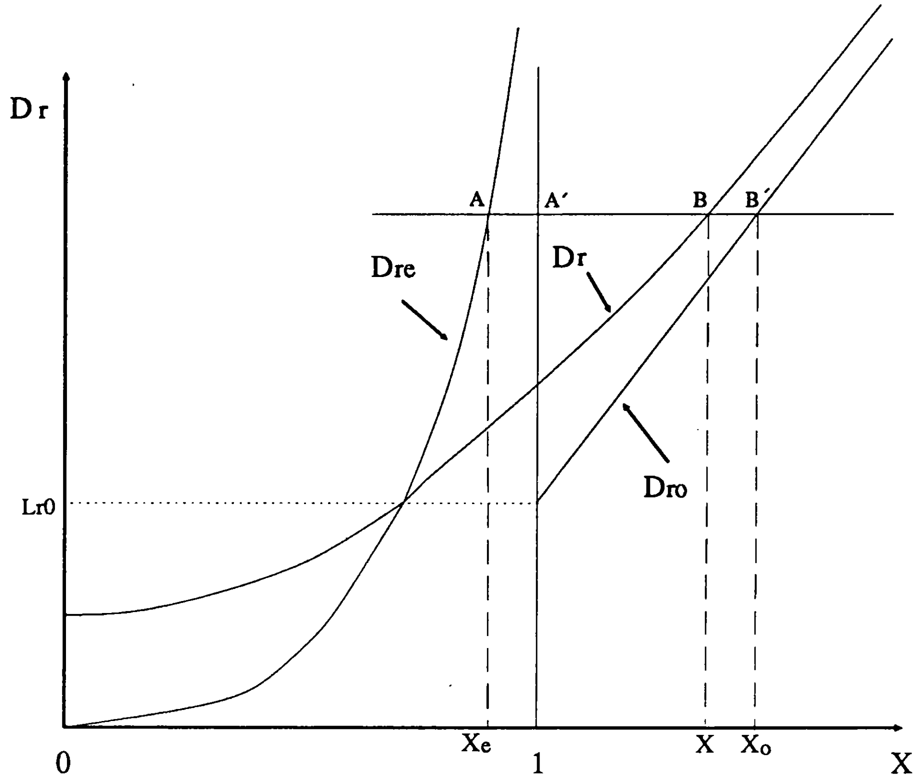


Fig 2.1 The Coordinate Transformation

Substituting X_e into (2.11a), and let $D_{re} = D_{ro} = D_r$, by solving a quadratic equation, we have

$$D_r = 0.5 \left(\sqrt{A^2 + B} - A \right) \tag{2.15}$$

where

$$A = \frac{(Qt)^2 + (4C - Qt)(qt + 2L_{r0})}{2(Qt - 2C)}$$

$$B = \frac{2C\,(qt + 2L_{r0})^2}{Qt - 2C}$$

However, this formula only gives an estimation for the random delay, and in the case of traffic signals, the uniform delay should also be considered. This thesis solves this problem in Chapter 4, where the idea of coordinate transformation is adopted to develop a new comprehensive formula which can be used for estimating the total delay (random + uniform).

§2.4 TRAFFIC SIGNAL SETTINGS FOR A SINGLE JUNCTION

§2.4.1 The Undersaturated Junctions

The calculation of signal timings, i.e. the cycle time and green times which yield satisfactory operating conditions, is determined by a selection of a performance index to represent the effect of the signal settings on the traffic conditions. For a junction that is undersaturated, the commonly accepted criterion is that of minimising the average total rate of delay to the traffic in all streams at this junction, since the delay expressions are relatively simple and easy to calculate, and the reduction in delay usually results in a reduction in fuel consumption, queue length and the probability of vehicular stops. Research on delay-minimising has been done, for example, by **Webster** (1958, — and **Cobbe**, 1966), **Miller** (1963, 1968) and **Allsop** (1971).

Webster (1958) used his two-term expression for average delay per vehicle to calculate the total rate of delay at the junction. He assumed the selection of representative stream for each stage, that is, the stream which has the highest flow ratio (y values) in each stage. To deduce an expression for the cycle time which gives the least average delay to all traffic, he differentiated the equation for the overall delay at the junction with respect to the cycle time, and found that the cycle time with the minimum delay could be represented approximately by

$$c_0 = \frac{1.5L + 5}{1 - y_1 - y_2 - \cdots - y_m} = \frac{1.5L + 5}{1 - Y} \text{ seconds} \qquad (2.16)$$

where y_1, y_2, $\cdots$, y_m are the maximum ratios of flow to saturation flow for stages 1, 2, $\cdots$, m, $Y = \Sigma y_i$ and L is the total lost time per cycle in seconds. This formula applies as stated only if each representative stream has green only in the stage it represents. In other cases it can still be used but there are further steps in the calculation.

The green times can also be decided by letting the ratio of the effective green times be equal to the ratio of the y values, i.e.

$$\frac{g_1}{g_2} = \frac{y_1}{y_2} \qquad (2.17)$$

This is the case for the two stage situation, and can be extended to 3 or more stage operation. In (2.17) g_1 and g_2 are the effective green times of stage 1 and 2 respectively.

Miller (1963, 1968) considered the case of a two-stage intersection, and derived the cycle time that yields the least total delay:

$$c = \frac{L + 2(I_i L/S_i)^{1/2}}{1 - (y_i/\pi_i)} \qquad (2.18)$$

where i is the direction that has the highest flow, and π_i denotes the proportion of the available green time allocated to stream i, i.e. $g_i = \pi_i(c-L)$, and is given by

$$\pi_i = \frac{(y_i I_i)^{1/2} + 1.2(y_i y_j)^{1/2}\{(y_i I_j)^{1/2} - (y_j I_i)^{1/2}\}}{(y_i I_i)^{1/2} + (y_j I_j)^{1/2}} \qquad (2.19)$$

where i=1 and j=2 or vice versa.

For three or more stages, the allocation of green time is found by solving a set of equations. This method is similar to **Webster's** in that it also requires the selection of a representative stream for each stage, and the two methods give very similar results.

Although **Webster's** method is quite straightforward and gives satisfactory results for simple junctions, it may just give the signal settings which would be no more than a reasonable approximation to the optimum when the junctions are more complicated. Both **Webster's** and **Miller's** methods are limited by the selection of a representative stream in each stage.

Allsop (1970, 1971) began the systematic application of optimization techniques to solve the problem. His method uses the same criterion of minimising the estimated average total rate of delay, but differs from previous methods in that it considers all the streams to the junction rather than only the representative streams so that it is easily applicable to complicated situations.

He shows that the optimisation problem can be solved by a convex programming approach. Mathematically, the problem can be expressed as:

$$\text{Minimise } D_e(\underline{\lambda}) = \sum_{j=1}^{n} \{\frac{1}{\lambda_0} f_j(\Lambda_j) + g_j(\Lambda_j)\} \qquad (2.20)$$

where $\underline{\lambda} = (\lambda_0, \lambda_1, \cdots, \lambda_m)$

$$f_j(\Lambda_j) = \frac{Lq_j(1-\Lambda_j)}{2(1-y_j)} \qquad (j=1, 2, \cdots, n)$$

$$g_j(\Lambda_j) = \frac{y_j^2}{2\Lambda_j(\Lambda_j-y_j)} \qquad (j=1, 2, \cdots, n)$$

$$\Lambda_j = \sum_{i=0}^{m} a_{ij}\lambda_i \qquad (j=1, 2, \cdots, n) \qquad (2.21)$$

and $0 \leq a_{0j} < 1$ $\qquad (j=1, 2, \cdots, n) \qquad (2.22)$

subject to the constraints

$$S(\underline{\lambda}) = \sum_{i=0}^{m} \lambda_i = 1 \qquad (i=1, 2, \cdots, m) \qquad (2.23)$$

$$\sum_{i=0}^{m} a_{ij}\lambda_i \geqslant y_j/P_j \qquad (j=1, 2, \cdots, n) \qquad (2.24)$$

where P_j is the maximum acceptable degree of saturation for stream j.

(capacity constraints)

$$\lambda_0 = k_0 \quad \text{or } \lambda_0 \geq k_0 \qquad \text{(cycle-time constraint)} \qquad (2.25)$$

and $\lambda_i \geq k_i\lambda_0$ $\qquad (i=1, 2, \cdots, m) \qquad (2.26)$

(minimum green constraints)

where $0 \leq k_i \leq 1$ $\qquad (i=1, 2, \cdots, m) \qquad (2.27)$

and if $k_0 \neq 0, \sum_{i=0}^{m} k_i \leq 1/k_0 - 1$ $\qquad (2.28)$

where m = the number of stages in the signal cycle

n = the number of streams

$\lambda_0 = L/c$

a_{0j} = proportion of the total lost time that is effectively green for stream j

$$\text{and} \qquad k_0 = \begin{cases} L/c_m & \text{if it is required that } c \leq c_m \\ L/c_s & \text{if it is required that } c = c_s \end{cases} \qquad (2.29)$$

where c_m = maximum cycle time

c_s = specified cycle time

$k_i = g_{im}/L$ $\qquad$ $(i=1, 2, \cdots, m)$

g_{im}= minimum effective green time for stage i $\qquad$ $(i=1, 2, \cdots, m)$

Then the vector $\underline{\lambda}$ will completely determine the signal settings, and the $\underline{\lambda}$ that minimises $D(\underline{\lambda})$ is the optimal solution, which can be obtained by a steep descent method. As a result of this method, a well known computer program SIGSET was developed (**Allsop** 1971b).

There is very little difference between the methods of **Webster**'s and **Allsop**'s when they are used for simple junctions, but **Allsop**'s method gives more reduction in delay than that of **Webster**'s when the junctions are complicated.

The stage-based modelling method by **Allsop** is used as the framework for this thesis.

The above introduced methods can be used for delay-minimising signal settings for an undersaturated junction. When it is possible that the junction becomes oversaturated, they cannot be used directly due to the deficiencies of their models. The next section §2.4.2 introduces methods that may be used for the potentially oversaturated junctions.

§2.4.2 Junctions That May Be Oversaturated

Although formulae like (2.15) are available for estimating the random delay per unit time for any X, different expressions for the uniform delay D_u have to be used when X<1 and X≥1. For example, when X < 1, the delay per unit time can be expressed as:

$$D = D_e = D_r + D_{ue} \tag{2.30a}$$

where D_r is given by (2.15), and D_{ue} by (2.7b).

When X ≥ 1, D becomes:

$$D = D_o = D_r + D_{uo} \tag{2.30b}$$

where D_r is still given by (15), but D_{uo} is given by (2.10).

The different expressions for D_u in the different range of X will cause difficulties when we minimise the delay for a junction at which some or all of the streams in the junction may be oversaturated, since in this case D_u is non-convex in Λ and non-differentiable (therefore $D = D_r + D_u$ is non-convex and non-differentiable), and analytical methods that have to use the derivatives such as $\partial D/\partial\Lambda$ will fail. This problem will be discussed further in §4.2 in the next Chapter.

Generally, the current methods of setting traffic signals for a junction that may be oversaturated can be divided into three categories:

(1) Methods that use the total rate of delay or total delay as the objective function, but when X ≥ 1, still use D_{ue} instead of D_{uo}, i.e. the traffic signal settings are obtained through minimising (2.30a) for any X, therefore give an inaccurate estimate of the uniform delay when X ≥ 1, since D_{uo} and D_{ue} are different. However, by such a treatment, the difficulties arising from non-convexity and non-differentiablity can be avoided, since D_{ue} becomes convex and differentiable.

(2) Methods that find the optimum traffic signal settings with respect to other objective functions rather than delay.

(3) Methods that use different objective functions for different X. i.e. when $X < 1$, minimising the delay, when $X \geq 1$, maximising the junction capacity.

It should be pointed out that some computer programs are capable of finding the optimum traffic signal settings with respect to different alternative objective functions, therefore they may belong to more than one of the three classes above.

Examples of (1) are the TRAFSIG method (**Reljic** 1988) and **SIGSIGN** (**Sang** and **Silcock** 1989). The latter is a phase-based program. The phase-based approach is reviewed in §2.5.

Examples of (2) are :

(a) Maximising the junction capacity, e.g. the SIGCAP method (**Allsop** 1972, 1976), the TRAFSIG method and the SIDRA-2 method (**Akcelik** and **Besley** 1984);

(b) Minimising the degree of saturation (**Ohno** and **Mine** 1973).

(c) Minimising the stop rate or fuel consumption, e.g. the TRAFSIG method.

An example of (3) is the program OSCADY (**Burrow** 1987).

The next part §2.5 gives a brief review on the phase-based approach. Although it is not going to be employed in establishing the framework for the research in this thesis, it does provide the possibility for further improvements, as a more flexible technique, and the new methods developed in this thesis are applicable to it.

§2.5. THE PHASE-BASED CONTROL METHOD

§2.5.1 Introduction

Almost all of the above-mentioned methods are stage-based. However, the development of modern microprocessor controllers has made it possible to achieve a new type of control — the phase-based control approach to meet the increasing complexity of signalised junctions. To illustrate the differences between the two methods, first we need to review some important terminology. A phase is the sequence of signal conditions given to one or more vehicle or pedestrian streams of traffic so that each stream allocated to the phase receives identical signal indications. A stage is a part of the signal cycle during which a particular set of phases gets green. Each stage is usually separated in the signal cycle by a stage intergreen period which is calculated from the clearance times required between streams losing green and streams gaining green. A stage sequence is the order in which the stages occur in a signal cycle.

In stage-based control methods, the interstage structure and sequence is determined as an initial step; only the cycle time and the green timings are regarded as variables, and calculated to optimise a certain objective function, e.g. delay, number of stops, or capacity. For each stage the calculation of the clearance times must be made for all the pairs of streams which could conflict, but the duration of the stage intergreen will usually be the maximum of the results. Even if a more complicated interstage structure is used, this has to be chosen manally and not as part of the optimisation process. Since the stage composition and sequence are externally fixed, the stage-based approach cannot assure a globally optimal control system design. Whereas in phase-based control, the phase-to-phase intergreens can be set directly in the signal controller rather than grouped into stages. Thus when changing the stages, the starts and ends of each phase can be staggered according to the intergreen time required between conflicting phases instead of all the phase greens starting and ending simultaneously. The stage sequence and interstage structure, together with the cycle time and green times, are all considered as variables, so a global optimum can be attained, at least in principle.

§2.5.2 Previous Work

The first important work was done by **Stoffers** (1968), who develops the idea of joint optimisation of stage composition, stage sequence, cycle time and green times. He defines the stage matrix, then subdivides the problem into three separate steps: (i) find all the stages that agree with the compatibility matrix; (ii) construct all maximal stage sequences which yield one green interval per cycle for each of these phases; (iii) find the optimal schedule for each possible stage sequence.

Tully (1977) continued **Stoffers**'s work in her PhD thesis. She develops the computer program SQGN which determines all possible stages and maximal stage sequences. Each such sequence and their subsequences need to be evaluated. However, when the number of sequences generated is big, evaluation is very difficult, and also in this case the computer cost is not negligible.

Tully and **Murchland** (1978) use the critical cycle time as an evaluation parameter to reduce the number of sequences. Only those sequences that have a good value of the critical time would be further explored, hence the computation time could be reduced to some extent. The above procedures give every phase only a single green interval, thus limiting their use. Improvements have been made to introduce double green facilities in a modified version of SQGN (**Heydecker** 1982, **Morton** 1987).

Improta and **Cantarella** (1984) propose an approach to the problem which allows all the control variables to be incorporated into a binary mixed-integer linear programming model. They use binary variables to represent the order in which each pair of incompatible groups receive green and the mixed-integer optimisation is performed by a Branch-and-Bound method.

Gallivan and **Heydecker** (1988) use the maximal sequences generated by SQGN as a basis to solve the optimisation problem. They consider the signal timings for each group directly instead of through the stages, and this makes it convenient to express the control variables. As a result, the optimization problems are either linear or convex, depending on the form of objective functions and the type of constraints, and can be solved by mathematical programming approach. However, the stage sequence is difficult to optimise when the number of sequences generated by SQGN is big.

MÖller (1987) developed an approach which minimises the number of constraints by considering the constraints on clearance times; some of those constraints can be omitted under practical conditions. The resulting optimization problems are linear for the minimum cycle time and maximum capacity, and a convex objective function can be used for the minimum delay. Finally, an event-based approach can be used in the case of coordinating traffic signals.

§2.5.3 The Variables and Constraints

In the phase-based control method, the main variables are the time during a cycle at which each phase first gets green and the duration of that green interval. The signal cycle is represented by the interval (0,c), where c is the cycle time. If N is the number of phases and M is the number of stages required to appear, then

θ_i = the proportion of the start of the green for phase i in the cycle time $(1 \le i \le N)$

ϕ_i = the proportion of the duration of the green for phase i in the cycle time $(1 \le i \le N)$

Ω_i = the proportion of the start of the minimum or core green for stage i in the cycle time, where stage i is the ith stage required to appear $(1 \le i \le M)$

The cycle time is represented indirectly by using the variable $\tau = 1/c$

By definition, the Variables θ_i and ϕ_i should satisfy the constraints:

$$0 \le \theta_i \qquad (1 \le i \le N) \qquad (2.31a)$$

and

$$0 \le \phi_i \le 1 \qquad (1 \le i \le N) \qquad (2.31b)$$

The optimisation of these variables is also subject to a number of practical engineering constraints for the sake of safety and efficiency. One important constraint is the intergreen time requirement to ensure that the start of phase i is at least one intergreen time later than the end of phase j. If $\theta_j+\phi_j+I_{ji}<1$, Then the constraint would be:

$$\theta_j + \phi_j + I_{ji} \le \theta_i \qquad (1 \le i,j \le N) \qquad (2.32)$$

and if $\theta_j+\phi_j+I_{ji} \ge 1$, then

$$\theta_j + \phi_j + I_{ji} - 1 \le \theta_i \qquad (1 \le i,j \le N) \qquad (2.33)$$

I_{ji} = the proportion of the intergreen time in the cycle time from phase j to phase i, which depends on the geometric layout of the junctions.

The reason for the correction of -1 in (2.33) is that both θ_i and θ_j are measured from the start of the same cycle in both (2.32) and (2.33), whereas the right hand side of (2.33) would, if the -1 were not substracted, relate to the next occurance of green for phase i.

In practice the cycle time can be either specified or subject to a maximum value. In the former case the cycle time should satisfy

$$\tau \leq 1/c_s \tag{2.34a}$$

and

$$\tau \geq 1/c_s \tag{2.34b}$$

In the latter case, however, the constraint becomes

$$\tau \geq 1/c_m \tag{2.35}$$

where

c_s = specified cycle time
c_m = maximum cycle time

Similarly, the green time for each phase is subject to the following constraint:

$$\phi_{im} \leq \phi_i \qquad (1 \leq i \leq N) \tag{2.36a}$$

if there is a minimum green time ϕ_{im} for phase i, and

$$\phi_{is} \leq \phi_i \leq \phi_{is} \qquad (1 \leq i \leq N) \tag{2.36b}$$

if the green time for phase i is specified as ϕ_{is}.

Further constraints are needed to specify that particular stages should appear in the solution with some minimum duration and that they should occur in a particular order. First define β_i as the minimum or core green for the ith stage required to appear ($1 \leq i \leq M$, M is the number of stages required to appear). If there is just one stage which is required to appear, then its core green is set to start at the origin; if there is more than one stage required to appear, then the core green of the first specified stage is set to start at the origin. For each phase in a required stage, there are two constraints. The two constraints need only be specified once for each phase, even when the phase appears in more than one required stage. Let F_i be the first stage required to appear in which phase i gets green and L_i the last. Firstly, for each phase appearing in such a stage, the phase green should start either before the start of the core green for the stage or at the same time:

$$\theta_i \le \Omega_{F_i} \qquad \text{where} \quad F_i \ne 1 \qquad (2.37a)$$

or

$$\theta_i \le 1 \qquad \text{where} \quad F_i = 1 \qquad (2.37b)$$

In addition, the stage core green for stage L_i must end either before the end of the phase green or at the same time. In other words, the start plus the duration of the stage core green for stage L_i should be less than or equal to the start plus the duration of the phase:

$$\theta_i + \phi_i \ge 1 + \beta_{L_i} \qquad \text{when } L_i = 1 \qquad (2.38a)$$

or

$$\theta_i + \phi_i \ge \Omega_{L_i} + \beta_{L_i} \qquad \text{when } L_i \ne 1, \; L_i \ge F_i \qquad (2.38b)$$

or

$$\theta_i + \phi_i \ge 1 + \Omega_{L_i} + \beta_{L_i} \qquad \text{when } L_i \ne 1, \; L_i < F_i \qquad (2.38c)$$

In the situation where there is no stage required to appear, however, the starts of the phase greens as calculated by solving optimisation problems are not all within the range 0 to 1 in the initial solution, and multiples of 1 should be subtracted to obtain the final solution. It is sufficient to apply (2.31a), and (2.37) and (2.38) are omitted.

The final constraint is the capacity constraint which is necessary to ensure that if possibile each traffic stream gets enough green time for traffic in the stream to pass the junction. Let

$$g(j) = \text{the index of the phase which controls stream } j \qquad (1 \le j \le n)$$

The capacity constraints are

$$q_j \le P_j[\phi_{g(j)} + e_j\tau]s_j \qquad (1 \le j \le n) \qquad (2.39a)$$

where, as usual

n = the number of streams at the junction

q_j = the mean arrival rate of vehicles in stream j $(1 \le j \le n)$

s_j = the saturation departure rate for stream j $(1 \le j \le n)$

e_j = the amount of extra effective green time that stream j has due to start and end lags caused by acceleration of vehicles $(1 \le j \le n)$

and P_j = the maximum acceptable degree of saturation in stream j $(1 \le j \le n)$

If the arrival rates in all the streams is multiplied by a common factor μ, then in order to accommodate the new flows, the following constraint must be satisfied:

$$\mu q_j \le P_j[\phi_{g(j)} + e_j\tau]s_j \qquad (1 \le j \le n) \qquad (2.39b)$$

§2.5.4 Optimisation of Signal Timings

The problem of applying the phase-based approach to optimal signal timings for a single junction has been discussed by **Heydecker** and **Dudgeon** (1987) and **Morton** and **Silcock** (1987). Optimisation is carried out for each relevant set of choices of the alternative constraints (2.32) and (2.33).

(a) Minimising the cycle time

This is especially important in the situation of linked traffic signals where all junctions operate on the same cycle time. The shortest cycle time can be obtained by solving the linear programming problem:

$$\text{Maximise } \tau = 1/c \qquad (2.40)$$

subject to constraints (2.31)-(2.39).

(b) Maximising the junction capacity

Here the objective function is the common factor by which the arrival rates in all of the streams can be increased before any of the capacity constraints is violated. This is also a linear programming problem:

$$\text{Maximise } \mu \qquad (2.41)$$

subject to the constraints (2.31)-(2.39).

(c) Minimising the junction delay

The estimated average total rate of delay for a junction can be expressed as:

$$D = \sum_{j=1}^{n} q_j d_j \tag{2.42}$$

where d_j is the mean delay for each stream. Using **Webster's** two term delay formula, d_j can be expressed as:

$$d_j = \frac{9}{10}\left\{\frac{s_j(1-\Lambda_j)^2}{2(s_j-q_j)\tau} + \frac{q_j}{2s_j\Lambda_j(\Lambda_j s_j-q_j)}\right\} \qquad (1\le j\le n)$$

where $\Lambda_j = e_j\tau + \phi_{g(j)} \qquad (1\le j\le n) \qquad (2.43)$

The optimisation problem can then be expressed as:

$$\text{Minimise} \quad D \tag{2.44}$$

subject to constraints (2.31)-(2.39), (2.42), (2.43).

§2.6. COMPUTER PROGRAMS FOR FIXED-TIME TRAFFIC SIGNALS

§2.6.1 Introduction

There are already many computer programs available for fixed-time traffic signal settings at an individual junction, some of which employ the comprehensive formulae that can model the situations where the junction is oversaturated. Widely acailable stage-based computer programs include: SIGSET (**Allsop** 1971b), SIGCAP(**Allsop** 1976), SIDRA-2(**Akcelik** & **Besley** 1984), SOAP84 (**Federal Highway Administration** 1985), OSCADY (**Burrow** 1987). Another such program that has been described for practioners is TRAFSIG (**Reljic** 1988). Since the work in this thesis is most closely related to the program OSCADY, a description of it is given in the next section §2.6.2. There are two widely available phase-based computer programs: SIGSIGN (**Sang** and **Silcock** 1989) and LINSIG (**Simmonite** 1985). Table 2.1

gives a summary of these programs. The method of **Improta** and **Cantarella** and of **Möller** have also been implemented in computer programs but these have not been publicisied among practioners.

It can be seen from Table 2.1 that these programs are very useful for setting traffic signals either for equilibrium conditions or for a certain period of time, with respect to some choices of objective functions.

§2.6.2 Program OSCADY

The program OSCADY (**Burrow** 1987) can model capacities, queues and delays at isolated signal-controlled junctions. It can model 3-arm junctions and 4-arm layouts controlled by traffic signals. OSCADY uses time-dependent queueing theory (**Kimber** and **Hollis** 1979) to model the growth and decay of queues. This is achieved by considering the modelled period as a sequence of short time segments. During each segment the demand flow, capacity and signal timings are considered constant. Usually each such segment is about 5 to 15 minutes long. The program then allows the calculation of the queue length at the end of each segment and the delay occurring during the segment.

The optimisation techniques employed to derive the best signal timings are based on methods developed by **Allsop** (1971, 1972). First, the optimisation routines seek to derive settings such that all traffic streams operate within their practical capacity. The optimiser uses linear programming techniques to obtain signal settings which maximise the reserve capacity of the junction. If the settings obtained in this way produce a zero or negative reserve capacity then the junction is oversaturated and the signal settings which maximise the reserve capacity will be adopted, and no further calculations are needed. On the other hand, if the reserve capacity for these settings is positive, the junction is operating within capacity. In this case the optimiser then derives delay minimising settings.

Table 2.1 Computer Programs For Setting Traffic Signals

PROGRAM	Objective Function	Method of Specifying The Junction Layout	Ability To Give Penalty To Queues	Ability to Model Periods of Time - Varying Demand
SIGSET	If $X<1$ Total Delay	Stream By Stream	No	No
SIGCAP	Capacity	Stream By Stream	No	No
SIGSIGN	Critical Cycle Time; Total Delay; Capacity	Stream By Stream	No	No
LINSIG	Junction Capacity	Stream By Stream	No	No
SIDRA-2	Stop rate; Junction Capacity; Critical Cycle Time	Lane By Lane	No	No
SOAP-84	Weighted Sum of Delay And Number of Stops	Stream By Stream	No	No
OSCADY	If $X<1$ Total Delay; If $X\geq1$ Junction Capacity	Lane By Lane	No	Yes
TRAFSIG	Total Delay; Total Number Of Stops; Total Person Delay, etc.	Stream By Stream	Yes	No

§2.7 CONCLUSION

Although the programs discussed above are very useful, they leave something to be desired. The following points set out the reasons for carrying out the work described in this thesis:

(a) Only the programs TRAFSIG and SIGSIGN can conduct delay minimisation when the junction is oversaturated, and in these programs the formula for estimating the uniform delay in such a case is still given by D_{uc} as expressed by (2.7b), which is correct only for undersaturated situations.

(b) Only OSCADY can deal with periods of time-varying demand, but the resulting signal settings are only local solutions since they are obtained through the optimisation for each single time period. Moreover, the program can only perform capacity maximisation for those periods that are oversaturated, rather than delay minimisation. In addition, it can only model 3-arm junctions or 4-arm layouts.

This thesis therefore aims to make improvements with respect to (a) and (b), by

(i) developing an new comprehensive delay formula that can be used for minimising total delay for an isolated junction for **a single time period** for the whole range of values of X, and

(ii) developing an optimisation method that can be used for minimising total delay for **periods of time-varying demand** for a general junction.

Ways of carrying out tasks (i) and (ii) will be described in Chapter 4 and Chapter 5 respectively.

CHAPTER 3. ESTIMATION OF QUEUE LENGTHS AND DELAYS AT TRAFFIC SIGNALS

§3.1 INTRODUCTION

As mentioned in the earlier chapters, there are many performance indices that can be used in the process of calculating traffic signal settings, e.g. delay, queue lengths, stops, fuel consumption, etc. However, it is common practice to use total rate of delay D, or the total delay W for the junction, over a period T as the criterion for optimising traffic signal settings. If the value of D is the average over the period T, then W can be expressed as:

$$W = DT = T \sum_{j=1}^{n} D_j \qquad (3.1)$$

where D_j is the delay per unit time for stream j. Hence with this calculation of D there is no difference between the signals settings when D or W is used as the objective function.

D_j can be expressed as:

$$D_j = D_{uj} + D_{rj} \qquad (3.2)$$

Where D_{uj} and D_{rj} are the uniform delay per unit time and random delay per unit time respectively.

Apart from delay estimation, queue length estimation is also necessary, since it is important to know the evolution of the queues, and when more than one time period is considered, queue length estimation becomes essential because the queues at the end of one time period will be the initial queue lengths for the subsequent time period.

Although some expressions for delay have already been given in the preceding chapter, further discussion is still necessary, since it is important to investigate these expressions in more detail before making the attempt to fulfill the tasks stated in §2.8.

The availability of various formulae developed so far for estimating queues and delays can be summarised in Table 3.1.

Table 3.1 Availability Of Formulae For Estimating Queues and Delays

		Equilibrium	Oversaturation	Comprehensive
		$(X < 1)$	$(X \geq 1)$	(for all X)
QUEUES & DELAYS	Uniform	D_{ue} L_{ue}	D_{uo} L_{uo}	Does not exist
	Random	D_{re} L_{re}	D_{ro} L_{ro}	*e.g. TRRL Sheared Delay Formula*
	Random + Uniform	*e.g. Webster's formula*	*Determinstic formulae* $D_{uo} + D_{ro}$; $L_{uo} + L_{ro}$;	Not yet available

In this chapter, the formulae whose availability is summarised in Table 3.1 are reviewed, and some methods of improving the accuracy of estimation are discussed. A method to fill the vacancy in the bottom right corner in Table 3.1 will be given in §4.4.2, where a comprehensive formula for total delay is developed.

§3.2 QUEUE LENGTH ESTIMATION

§3.2.1 Random Queue Length Estimation

1. A Commonly Accepted Geometrical Approximation

When $X < 1$, and arrivals are Poisson, the equilibrium expression is:

$$L_{re} = \frac{CX^2}{1 - X} \tag{3.3}$$

However, when $X \geq 1$, the determinstic expression is:

$$L_{ro} = L_{r0} + (q-Q)t \tag{3.4}$$

According to the coordinate transformation method as described by **Kimber** and **Hollis** (1979), the random queue length for a stream can be expressed by the following comprehensive formula:

$$L_r(t) = 0.5(\sqrt{A^2+B} - A) \tag{3.5a}$$

Where

$$A=\frac{(1-X)(Qt)^2-QtL_{r0}+2C(L_{r0}+XQt)}{Qt-C} = \frac{(Qt)^2+(2C-Qt)(XQt+L_{r0})}{Qt-C}$$

$$B=\frac{4C(L_{r0}+XQt)^2}{Qt-C}$$

However, since the denominators of A and B contain (Qt-C), which becomes zero when t=C/Q, this expression must be modified in numerical calculation, so that (Qt-C) is removed from the denominator of the expression for $L_r(t)$. This can be achieved by rearranging (3.5a):

$$L_r(t) = 0.5(\sqrt{A^2+B} - A) = \frac{0.5B}{\sqrt{A^2+B} + A}$$

$$= \frac{2C(XQt+L_{r0})^2}{(Qt)^2+(2C-Qt)(XQt+L_{r0})+Qt\sqrt{[(X-1)Qt+L_{r0}]^2+4C(XQt+L_{r0})}} \tag{3.5b}$$

When $L_{r0}\neq0$, expression (3.5b) can normally be used to evaluate $L_r(t)$, for all $t\geq0$, in which case the denominator of (3.5b) is always greater than zero. However, when $L_{r0}=0$, (3.5b) has to be modified further so that $L_r(t)$ can be evaluated for very small or zero t.

When $L_{r0}=0$, expression (3.5b) for $L_r(t)$ can be rearranged as:

$$L_r(t) = \frac{2CX^2Qt}{(1-X)Qt+2CX+\sqrt{[(1-X)Qt]^2+4CXQt}} \tag{3.5c}$$

From (3.5c) it is clear that for all $t\geq0$, $L_r(t)$ can be evaluated since the denominator is always positive. When t=0, $L_r(0)=0=L_{r0}$.

2. Use of The Formula In Practice

A formula such as (3.5) for estimating the random queue length is approximate and not transitive with respect to time. For example, if (3.5a) is regarded as a function of both L_{r0} and t, i.e.

$$L_r = L_r(L_{r0},\ t)$$

Then the following relationship

$$L_r(L_{r0},t_1) = L_r(L_r(L_{r0},t_2),t_1-t_2) \qquad (0<t_2<t_1) \tag{3.5d}$$

does not hold if L_r is calculated by (3.5a). However, the property expressed by (3.5d) must be true for the queue itself and therefore should hold for a formula to be used in practice. To get such a formula, the time origin can be shifted so that the expression (3.5a) is transitive with respect to time in any particular application. The resulting method as used in OSCADY is introduced below. To illustrate such a method, some numerical calculations are made. The notations used in the graphs are:

L_r(DIRECT): L_r given directly by (3.5).
L_r(OSCADY): L_r given by the OSCADY formulae introduced below.

And in all the examples in this chapter, the arrival rate for the example stream is $q = 1500$ pcu/hour.

The OSCADY method is: define the function

$$L(t,X,Q,C) = \frac{2CX^2Qt}{(1-X)Qt+2CX+\sqrt{[(1-X)Qt]^2+4CXQt}} \tag{3.6}$$

Obviously (3.6) is a special case of (3.5a) when $L_{r0} = 0$.

(1) If $X \geq 1$, or $X<1$ and $0 \leq L_{r0} < L_{rc}$, then

$$L_r(t)=L(t+t_0,X,Q,C) \tag{3.7a}$$

$$t_0 = \frac{L_{r0}\left(L_{r0}+2CX+\sqrt{L_{r0}^2 \quad +4CL_{r0}}\right)}{2Q(CX^2 + L_{r0}X-L_{r0})}$$

is chosen so that a queue growing according to (3.6) from a length of zero at time $t=-t_0$ would on average have length L_{r0} at time $t=0$.

Example 1a: $X = 1.10$, $L_{r0} = 15.0$ pcu, $T = 10$ minutes, $C=0.6$ (Fig 3.1a).
Example 1b: $X = 0.95$, $L_{r0} = 5.0$ pcu, $T =20$ minutes, $C=0.6$, $L_{re}=10.8$ (Fig 3.1b).

(2) If $X<1$ and $L_{r0}=L_{re}$, then

$$L_r(t)=L_{re} \tag{3.7b}$$

Example 2: $X = 0.95$, $L_{r0} = L_{re} = 10.83$ pcu, $T = 10$ minutes, $C=0.6$ (Fig 3.2)

(3) If $X<1$ and $L_{re}<L_{r0}\leq 2L_{re}$, then

$$L_r(t)=2L_{re}-L(t+t_0,X,Q,C) \tag{3.7c}$$

where

$$t_0 = \frac{L^*_{r0}\left(L^*_{r0} \quad +2CX+\sqrt{(L^*_{r0})^2+4CL^*_{r0}}\right)}{2Q(CX^2+L^*_{r0}X-L^*_{r0})}$$

$$L^*_{r0} = 2L_{re} - L_{r0}$$

The resulting graph of L_r decreasing from L_{r0} is the reflection in $L_r=L_{re}$ of the graph of L_r increasing from $2L_{re}-L_{r0}$ that would be given by (3.7a).

Example 3: $X = 0.95$, $L_{r0} = 15.0 > L_{re}=10.83$ pcu, $T = 30$ minutes, $C=0.6$ (Fig 3.3).

(4) If $X<1$ and $L_{r0}>2L_{re}$, then suppose that L_{r0} was previously an equilibrium random queue length under the degree of saturation $X_0 = L_{re}^{-1}(L_{r0})$, i.e.

$$L_{r0} = \frac{CX_0^{\,2}}{1 - X_0}$$

Comparison of Formulae for Lr

X = 1.10, Lr0=15.0, T=10 minutes, C=0.6

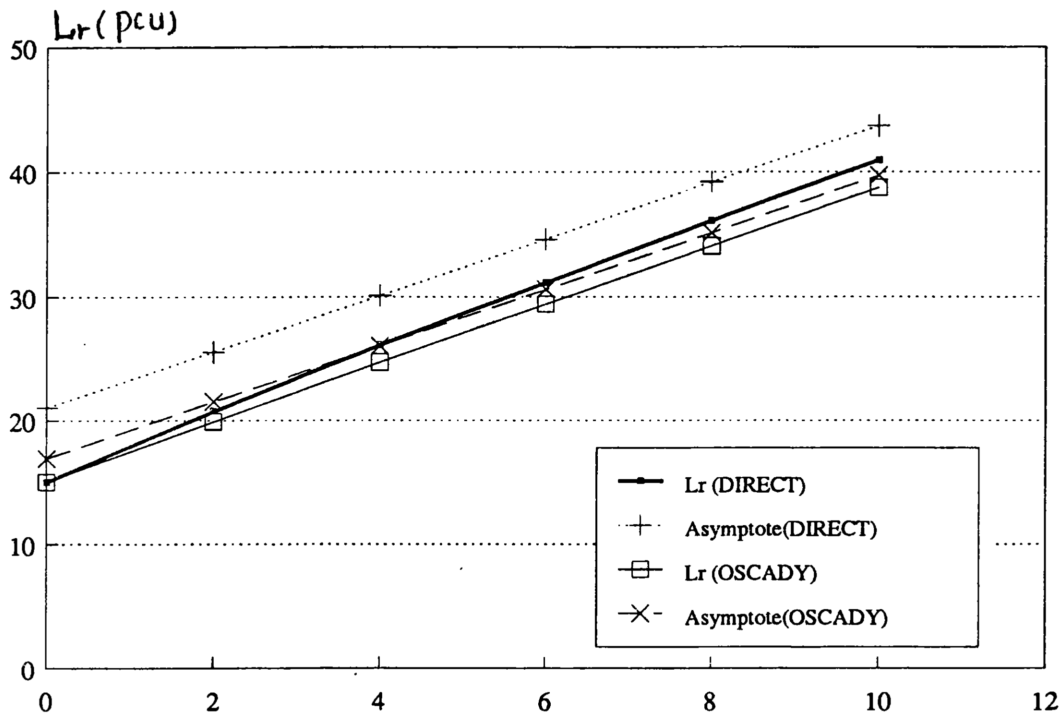


Fig 3.1a

Comparison of Formulae for Lr

X = 0.95, Lr0=5.0, T=20 minutes, C=0.6

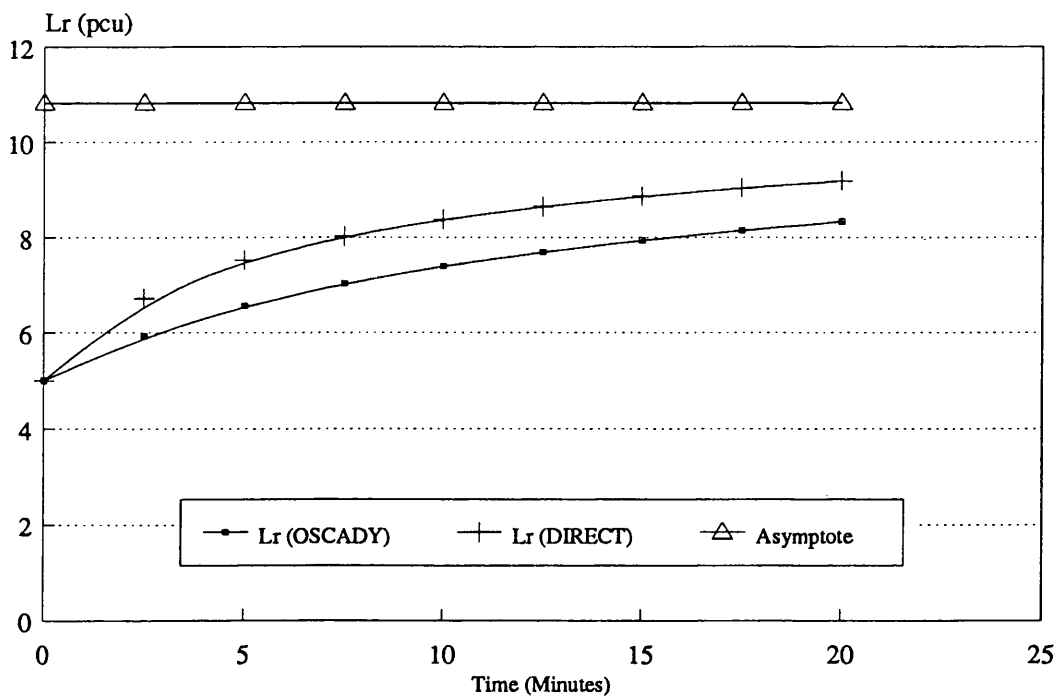


Fig 3.1b

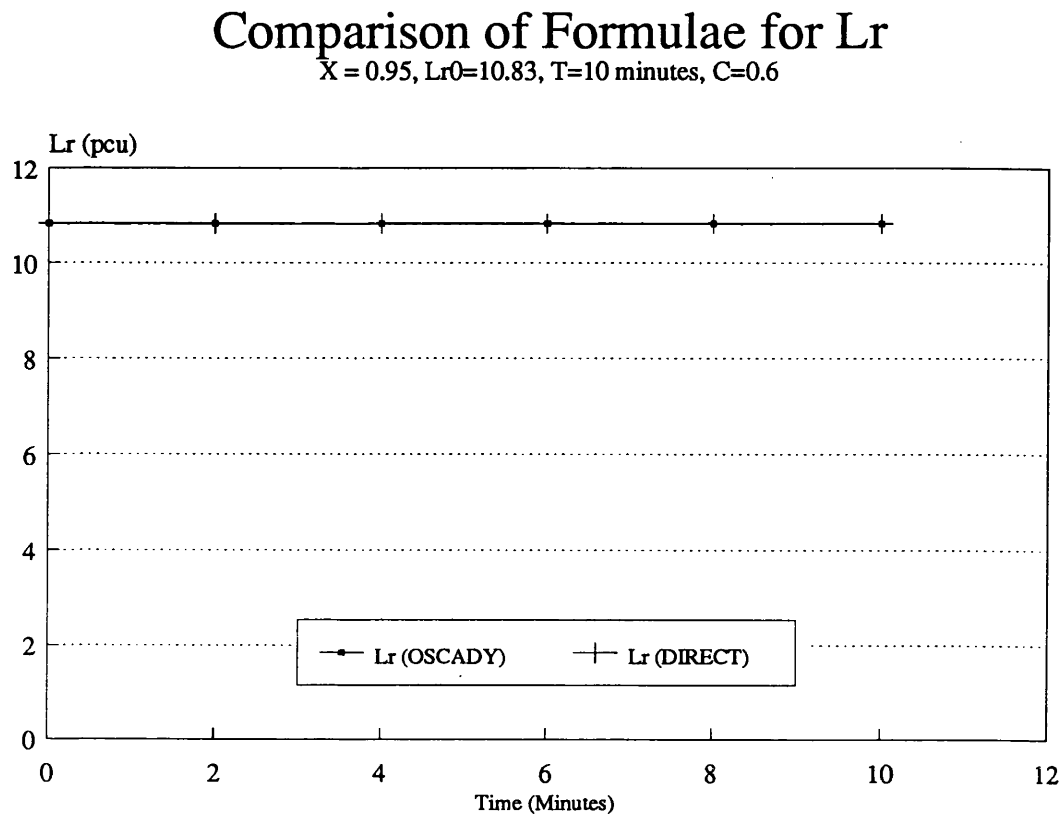

Comparison of Formulae for Lr
X = 0.95, Lr0=10.83, T=10 minutes, C=0.6
Lr (pcu)
12
10
8
6
4
2
0
Lr (OSCADY)
Lr (DIRECT)
0
2
4
6
8
10
12
Time (Minutes)

Fig 3.2

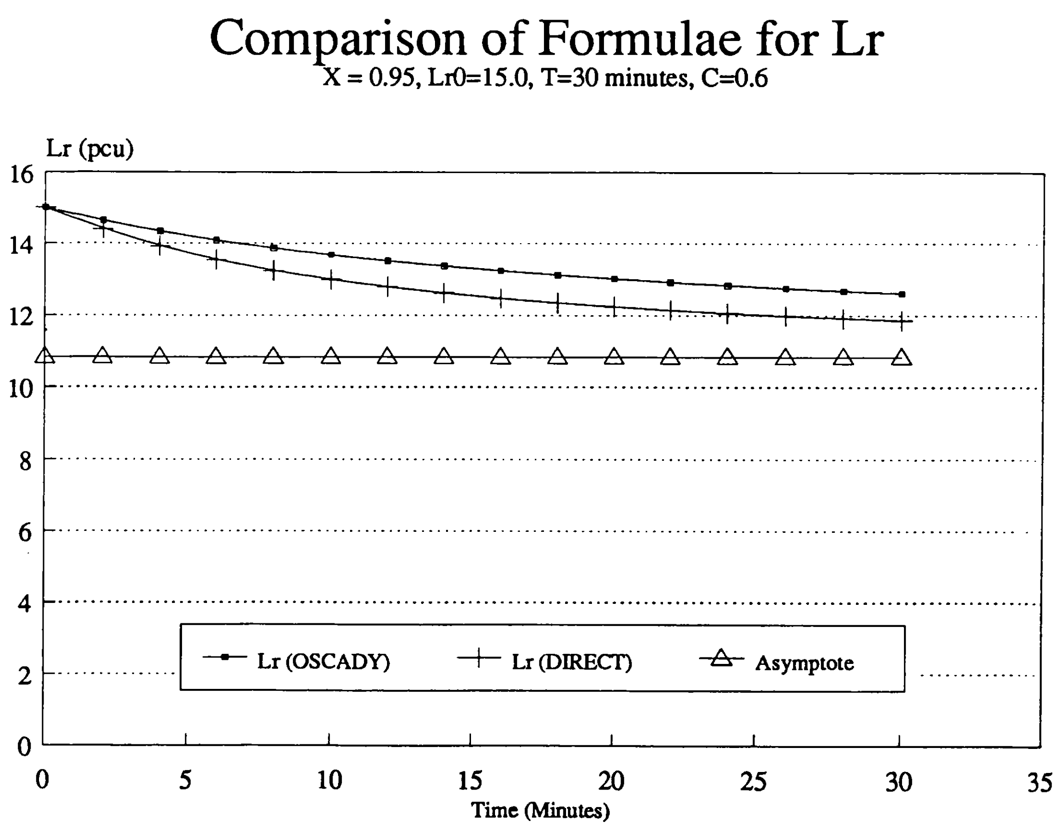

Comparison of Formulae for Lr
X = 0.95, Lr0=15.0, T=30 minutes, C=0.6
Lr (pcu)
16
14
12
10
8
6
4
2
0
Lr (OSCADY)
Lr (DIRECT)
Asymptote
0
5
10
15
20
25
30
35
Time (Minutes)

Fig 3.3

Hence $X_0 = \dfrac{\sqrt{L_{r0}^2+4CL_{r0}} - L_{r0}}{2C}$

Then starting from L_{r0}, L_r is considered as decreasing at a constant rate $Q(X_0-X)$ until $L_r = 2L_{re}$ after time t_c, after which L_r is considered as decreasing from $2L_{re}$ according to (3.7c).

Let $2L_{re}-L_{r0} = -\,Q(X_0-X)t_c$, then t_c can be given by

$$t_c = \frac{2C(2L_{re} - L_{r0})}{Q\left(2CX+L_{r0} - \sqrt{L_{r0}^2 + 4CL_{r0}}\right)}$$

Then

$$L_r(t)=\begin{cases} L_{r0} + Q\left\{X- \dfrac{\sqrt{L_{r0}^2 + 4CL_{r0}} - L_{r0}}{2C}\right\}t, & t \le t_c \\ \\ 2L_{re} - L(t-t_c, X, Q, C), & t > t_c \end{cases} \qquad (3.7d)$$

Example 4: $X=0.7$, $L_{r0} = 100.0 > 2L_{re} = 1.96$ pcu, $T = 30$minutes, $C=0.6$ (Fig 3.4).

According to **Kimber** and **Hollis** (1979), comparison of the random queue lengths given by (3.5a) and (3.7) with exact numerical calculations suggests that method as used in OSCADY does improve the accuracy of the geometric approximation. However, in case (4), the treatment introduced above is unrealistic, since the assumption that when $t<t_c$, L_r decreases at a rate $Q(X_0-X)$ will overestimate L_r. This is because according to the definition of capacity Q, L_r should on average decrease at a constant rate $Q(1-X) > Q(X_0-X)$ until the queues first clears, which will not usually happen before t_c. QX_0 is the average discharge rate over long periods in equilibrium, during which the queue has time to be empty for the average proportion $(1-X_0)$ of the period considered.

Although the two curves for L_r(DIRECT) and L_r(OSCADY) are different except in the rare case $L_{r0}=L_{re}$, they are all approximate and the differences are probably of limited importance provided that the same set of expressions is used for all streams and junctions being considered on any analysis or comparison between alternatives. For this reason and for the sake of comparability with other work, the formula used in OSCADY is retained for estimating L_r in this thesis.

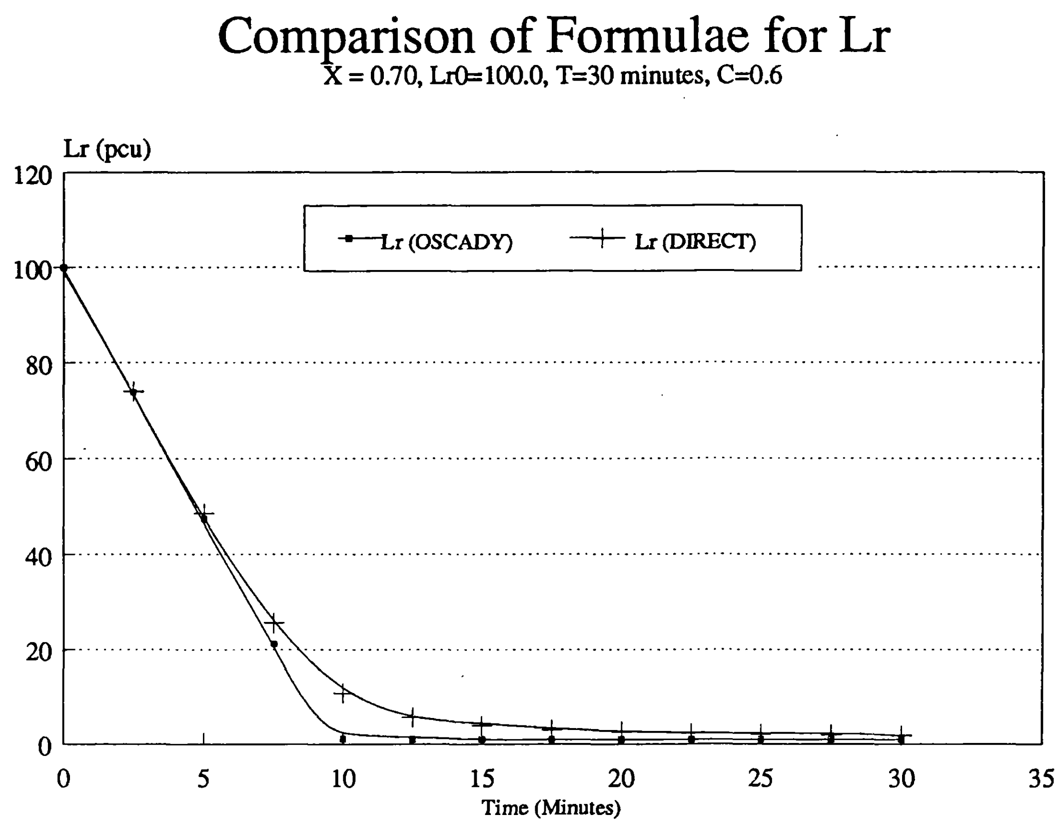


Fig 3.4

§3.2.2 Uniform Queue Length Estimation

The uniform queue and delay are the queue and delay that would occur due to the alternation of red and green at traffic signals if the traffic arrived at a uniform rate. During the red time a queue forms and this queue discharges at a constant rate s during the effective green time g. The **uniform queue length** L_u is then defined as **the average uniform queue length over T**, where T is the length of the time period considered.

1. Commonly Accepted Expressions

When $X < 1$, an equilibrium exists for the random queue length, and the average uniform queue length, L_u, can be expressed by:

$$L_u = L_{ue} = \frac{qc(1-\Lambda)^2}{2(1-\Lambda X)} \qquad (3.8)$$

This is the uniform delay term in **Webster's** formula, which can be illustrated by Fig 3.5a, in which the shaded area contributes to the

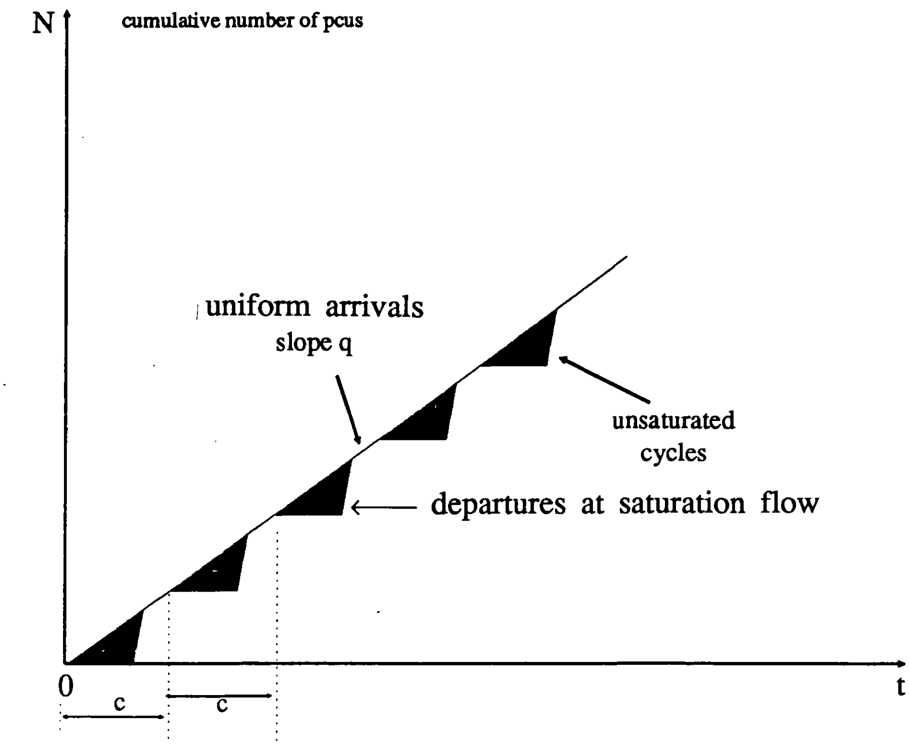


Fig 3.5a The approximate calculation of the uniform delay in the case of a equilibrium queue

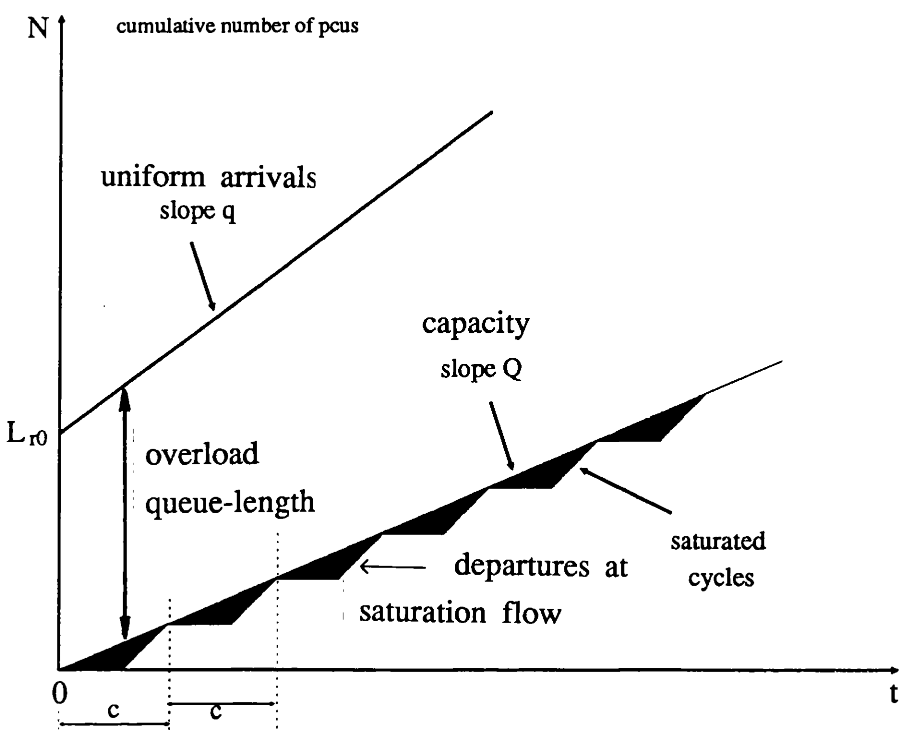


Fig 3.5b The calculation of the uniform delay in the case of a increasing queue

uniform queue length, and (3.8) can be obtained by calculating the average uniform queue length over a certain time period.

However, when $X \geq 1$, L_u can be expressed as:

$$L_u = L_{uo} = \frac{Qc(1-\Lambda)}{2} \tag{3.9}$$

This can be illustrated by Fig 3.5b, in which the shaded area contributes to the uniform queue length, and (3.9) can be obtained by calculating the average uniform queue length over a certain time period.

2. Calculating L_{uc} For More Accuracy When $X < 1$

Although (3.8) is widely accepted as the expression for estimating the uniform queue length when $X < 1$, it is only a rough formula, since it is based on the assumption that the random queue length L_r has reached the equilibrium, i.e. $L_r = L_{re}$. However, this assumption ignores the fact that if the initial random queue length L_{r0} is much greater than L_{re}, it will take some time, t_c, for L_{r0} to approach L_{re} (Fig 3.5c), therefore during this period L_u will be different from L_{uc}. Hence (3.8) is only suitable to use when L_{r0} is not much greater than L_{re}.

Therefore, for more accurate estimation of L_u, L_u will have different expressions when the degree of saturation X and the initial queue length L_{r0} are within different ranges.

(1) If $X \geq 1$

$$L_u = L_{uo} \tag{3.10a}$$

Where L_{uo} is given by (3.9).

(2) If $X<1$, there are two possibilities.

(i) If $L_{r0} \leq L_{re}$ (criterion (a))

then

$$L_u = L_{ue} \tag{3.10b}$$

where L_{ue} is given by (3.8).

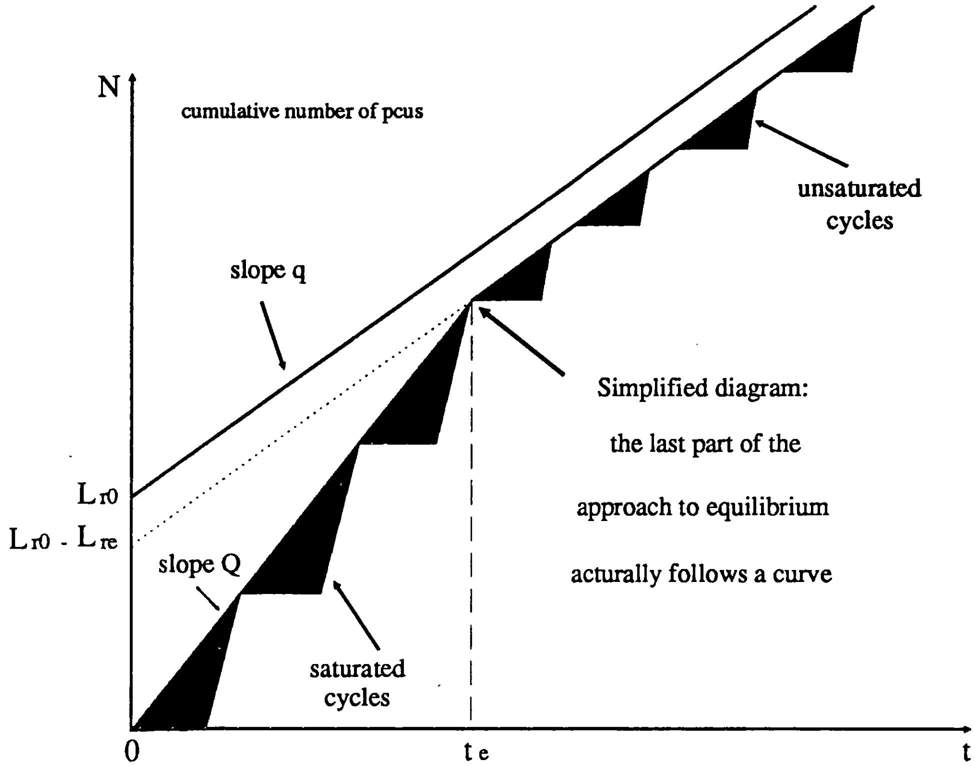


Fig 3.5c The calculation of the uniform delay in the case of a queue decreasing towards equilibrium

(ii) If $L_{r0}>L_{re}$ (criterion (b))

let $t_c = \dfrac{L_{r0} - L_{re}}{Q(1-X)}$ (Fig 3.5c), as an approximation to the time at which L_r approaches L_{re}.

If $t_c \geq T$ then

$$L_u = L_{uo} \tag{3.10c}$$

If $t_c<T$ then

$$L_u = \frac{Qc(1-\Lambda)}{2T}\left\{t_c + \frac{X(1-\Lambda)(T-t_c)}{1-\Lambda X}\right\} \tag{3.10d}$$

Criteria (a) and (b) proposed here are different from those used in OSCADY. In OSCADY, when $X < 1$, whether to use (3.10b) or (3.10c) or (3.10d) is judged by examining the degree of saturation of the stream in the previous period, rather than by examining the difference between L_{r0} and L_{re}, i.e.

when $X<1$, criterion (a) becomes:

(a′) If the stream was undersaturated during the previous period;

And criterion (b) becomes:

(b′) If the stream was oversaturated during the previous period.

In most cases this makes no difference between the two methods of choosing the expression for L_u, but the OSCADY method can sometimes be misleading. Even if the stream was undersaturated during the previous period, it does not necessarily mean that in this period $L_{r0} < L_{re}$. If there was a very long random queue at the beginning of the previous period, then this queue might not be able to be cleared to its equilibrium length before the end of that period. And even if it was nearly at its equilibrium, this equilibrium may be bigger than that in the current period.

Suppose that $L_{r0}>L_{re}$, then L_u should be estimated by (3.10c) or (3.10d). Let this be L_{u1}. However, since in the previous period the stream was undersaturated, then OSCADY method estimates L_u by $L_{u2} = L_{ue}$. Hence the absolute error in L_{u2} is:

$$\delta L_u = L_{u2} - L_{u1} = \begin{cases} L_{uc}-L_{uo} = \left(\dfrac{X-1}{1-\Lambda X}\right) L_{uo} & T \le t_e \\[2ex] \dfrac{(L_{uc}-L_{uo})t_e}{T} = \dfrac{(X-1)t_e}{(1-\Lambda X)T} L_{uo} & T > t_e \end{cases}$$

The error in L_{u2} relative to L_{u1} is:

$$100 \cdot \frac{\delta L_u}{L_{u1}}\% = \begin{cases} 100 \cdot \dfrac{X-1}{1-\Lambda X}\% & T \le t_e \\[2ex] \dfrac{100(X-1)t_e}{X(T-t_e)-\Lambda XT+t_e}\% & T > t_e \end{cases}$$

Example 1: in period 1, $X_1=0.95$, $L_{r01}=100$ pcu, $T_1=30$ minutes, at the end of T_1 the queue length is 65.2 pcu. In period 2, $X_2=0.7$, $L_{r02}=65.2$ pcu, $T_2=10$ minutes. Hence although during T_1 the stream was undersaturated, in T_2 it still holds that $L_{r02} > L_{re2} = 0.98$ pcu. The errors in the estimate given by the OSCADY method can then be seen in Table 3.2.

Table 3.2 Errors in L_{u2} compared with L_{u1}: Example 1

Cycle Time (seconds)	Absolute Error δL_u (pcu)	Relative Error $100\delta L_u/L_{u1}$(per cent)
20	-0.80	-35.62
40	-1.59	-35.62
60	-2.37	-35.62
80	-3.18	-35.62
100	-3.98	-35.62
120	-4.77	-35.62

Example 2: in period 1, X_1=0.95, L_{r01}=0.2 pcu, T_1=30 minutes, at the end of T_1 the queue length is 8.7 pcu. In period 2, X_2=0.7, L_{r02}=8.7 pcu, T_2=10 minutes. Hence although during T_1 the stream was undersaturated, in T_2 it still holds that $L_{r02} > L_{re2}$ = 0.98 pcu. The errors in the estimate given by the OSCADY method can then be seen in Table 3.3.

Table 3.3 Errors in L_{u2} compared with L_{u1}: Example 2

Cycle Time (seconds)	Absolute Error δL_u (pcu)	Relative Error $100\delta L_u/L_{u1}$(per cent)
20	-0.10	-6.22
40	-0.19	-6.22
60	-0.29	-6.22
80	-0.38	-6.22
100	-0.48	-6.22
120	-0.57	-6.22

Therefore whether the initial random queue length L_{r0} is bigger than the equilibrium random queue length L_{re} in the current period cannot simply be deduced by just examining the degree of saturation in the previous period. Instead, it should be judged by the direct calculation of L_{r0}, which usually has to be done anyway when successive periods with different levels of demand need to be modelled.

Therefore, when estimating the uniform delay, modifications to the formulae used in OSCADY are desirable, i.e. criteria (a′) and (b′) should be replaced by criteria (a) and (b). However, in many cases the two methods will give the same results, and for the sake of comparability with other work the OSCADY method is used in this thesis.

§3.3 DELAY ESTIMATION

Theoretically, the delay per unit time during a time period (0, t), can be obtained by the definition:

$$D_I(t) = \frac{1}{t}\int_0^t L(\tau)d\tau \tag{3.11a}$$

Where the subscript I stands for 'integral', and $L(\tau)$ can be one of the following:

$L_r(\tau)$;
$L_u(\tau)$;
$L_r(\tau)+L_u(\tau)$

when the corresponding D(t) needs to be calculated.

§3.3.1 Random Delay Estimation

The steady-state and determinstic expressions for the random delay per unit time are given by (2.11a) and (2.11b) in Chapter 2. The comprehensive formula can also be obtained. According to (3.11a), $D_r(t)$ can be calculated by:

$$D_{rI}(t) = \frac{1}{t}\int_0^t L_r(\tau)d\tau \tag{3.11b}$$

However, the direct calculation of D_r by this method would be complicated and slow, so that an approximate method is useful. Two approaches are available, which are introduced as follows:

(1) TRRL method – The Sheared Delay Formula

Kimber and **Hollis** (1979) derive a formula for estimating $D_r(t)$ (known as the Sheared Delay Formula), which was introduced as formula (2.15) in Chapter 2. It is rewritten again here as $D_{rs}(t)$, where the subscript s stands for 'sheared':

$$D_{rs}(t)=0.5(\sqrt{E^2+F}-E) \tag{3.12}$$

Where

$$E=\frac{(Qt)^2+(4C-Qt)(XQt+2L_{r0})}{2(Qt-2C)}$$

$$F=\frac{2C(XQt+2L_{r0})^2}{Qt-2C}$$

(2) Choice Of time-origin to make estimates transtive with respect to time

By methods involving choice of time-origin similar to those used for the estimation of the random queue length $L_r(t)$ when $L_{r0} \neq 0$, improvement can be made on $D_{rs}(t)$. The resulting formula is referred to as $D_{r0}(t)$, where the subscript 0 stands for 'time origin'. The first step is to define the function

$$D(t,X,Q,C)=\frac{2CX^2Qt}{(1-X)Qt+4CX+\sqrt{[(1-X)Qt]^2+8CXQt}}$$

(1) If $X\geq1$, or $X<1$ and $0\leq L_{r0}<L_{re}$, then

$$D_{r0}(t)=D(t+t_0,X,Q,C) \tag{3.13a}$$

where

$$t_0=\frac{L_{r0}\left(L_{r0}+2CX+\sqrt{L_{r0}^2+4CL_{r0}}\right)}{Q(CX^2+L_{r0}X-L_{r0})}$$

(2) If $X<1$ and $L_{r0}=L_{re}$, then

$$D_{r0}(t)=L_{re} \tag{3.13b}$$

(3) If $X<1$ and $L_{re}<L_{r0}\leq 2L_{re}$, then

$$D_{r0}(t)=2L_{re} - D(t+t_0,X,Q,C) \tag{3.13c}$$

where

$$t_0=\frac{L^*_{r0}\left(L^*_{r0}+2CX+\sqrt{(L^*_{r0})^2+4CL^*_{r0}}\right)}{Q(CX^2+L^*_{r0}X-L^*_{r0})}$$

$$L^*_{r0} = 2L_{re} - L_{r0}$$

(4)If $X<1$ and $L_{r0} > 2L_{re}$, then

$$D_{r0}(t) = \begin{cases} L_{r0}+0.5Q\left\{X-\dfrac{\sqrt{L^2_{r0}+4CL_{r0}}-L_{r0}}{2C}\right\}t & t\leq t_c \\ \dfrac{(L_{r0}+2L_{re})t_c}{2t}+[2L_{rc}-D(t-t_c,X,Q,C)]\dfrac{t-t_c}{t}, & t>t_c \end{cases} \tag{3.13d}$$

where

$$t_c = \frac{2C(2L_{re}-L_{r0})}{Q\left(2CX+L_{r0}-\sqrt{L^2_{r0}+4CL_{r0}}\right)}$$

To choose which of these two methods should be the one to be used as an alternative to D_{rI} for estimating random delay when accuracy is concerned, some numerical calculation in terms of the errors of these methods are necessary, and the one that gives smaller error should be employed. In the following examples, the data are:

$q=1500$ pcu/hour, $L_{r0}=5.0$ pcu, $T=30$ minutes, and $C=0.6$

In Fig 3.6-3.8, $X=1.2$, 0.95 and 0.7 respectively and the following notation is used to illustrate the results:

D_{rI}=The Delay calculated through the integration of $L_r(t)$ by using the *Romberg Algorithm* (**Burden** and **Faires** 1985), which is used for approximating the integral of a function f(x) over an interval [a,b] within a specified tolerance. The trapezoid rule is used to generate a preliminary approximation, and Richardson extrapolation (**Burden** and **Faires** 1985) is subsequently used to improve the approximation. Extrapolation continues until the fractional difference between successive approximations of the integral is less than the tolerance. D_{rI} is actually calculated by calling a standard Turbo PASCAL subroutine based on the above algorithm.

D_{rs}=The delay calculated by TRRL method;

D_{r0}=The delay calculated by the choice of time-origin.

The relative errors are given by:

$$RD_{rs} = \frac{100(D_{rs}-D_{rI})}{D_{rI}} \text{ per cent}$$

$$RD_{r0} = \frac{100(D_{r0}-D_{rI})}{D_{rI}} \text{ per cent}$$

The results show that when $L_{r0} \neq 0$, D_{r0} is closer to D_{rI} than D_{rs}. Hence when accuracy becomes important, a time-origin should be chosen to estimate the random delay per unit time.

However, in the program OSCADY, D_{rs} is used for estimating the random delay. For the sake of comparability with other work, D_{rs} is used in this thesis for D_r, although D_{r0} can give more accurate results when $L_{r0} \neq 0$.

§3.3.2 Uniform Delay Estimation

According to (3.11a), the average uniform delay per unit time over (0,t) can be calculated by:

$$D_{uI}(t) = \frac{1}{t} \int_0^t L_u(\tau)d\tau \tag{3.11c}$$

Referring back to §3.2.2, it follows by definition that numerically, the uniform delay per unit time and the uniform queue length for each stream are equal for the period (0,t), since by definition, L_u is the average uniform queue length over that period.

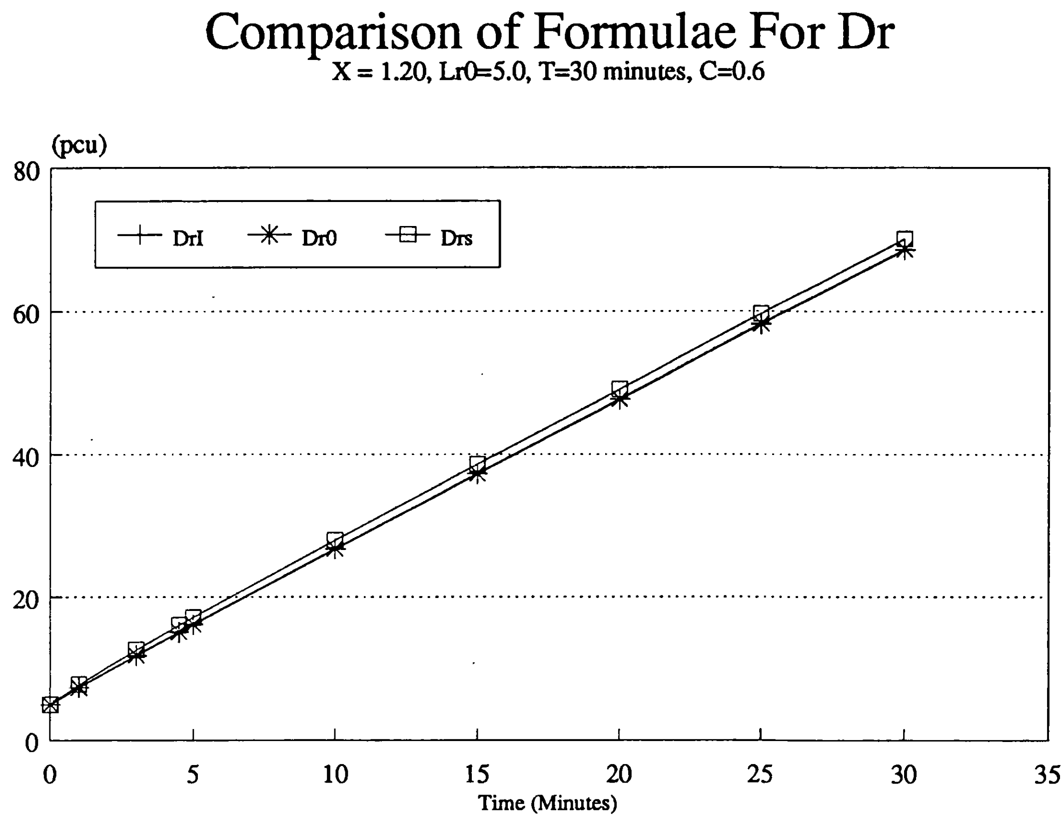


Fig 3.6a

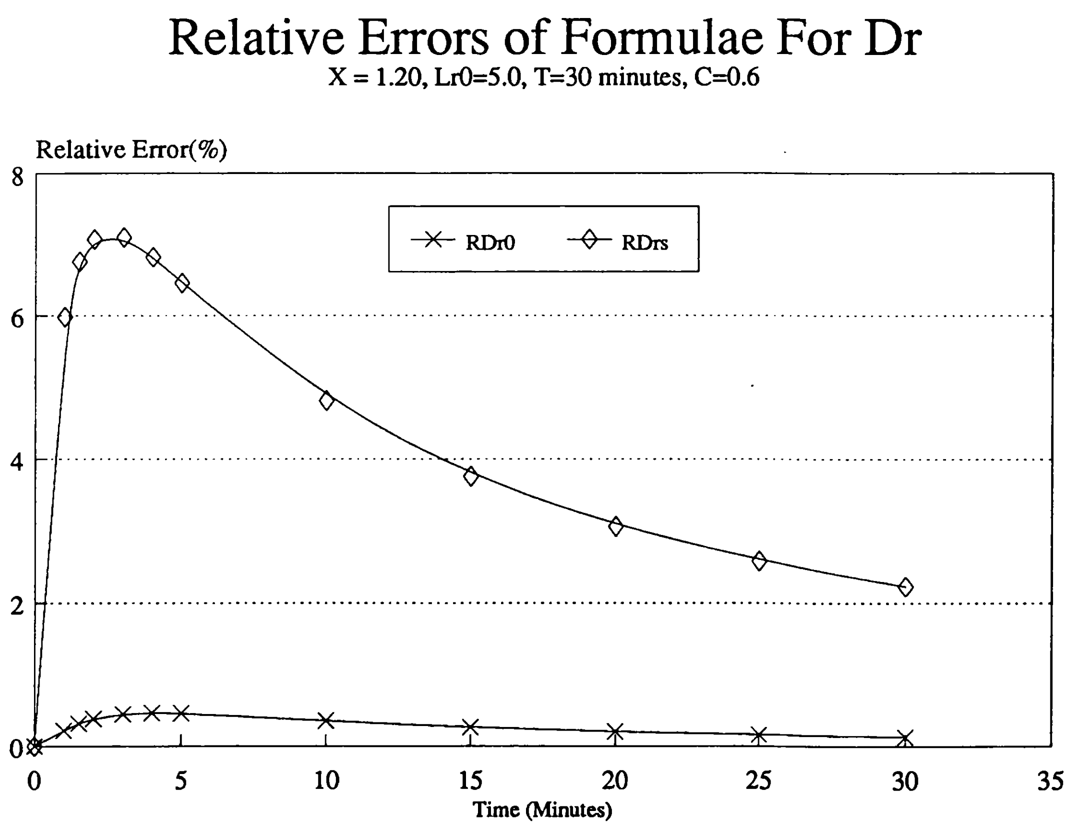


Fig 3.6b

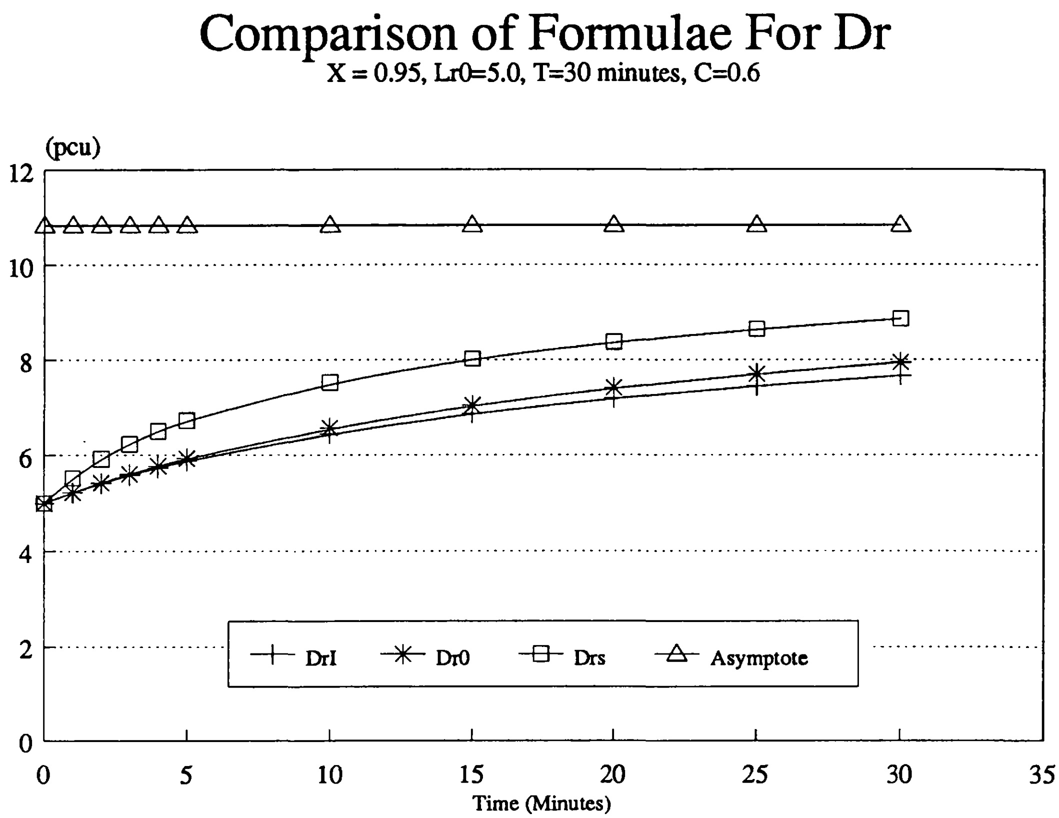


Fig 3.7a

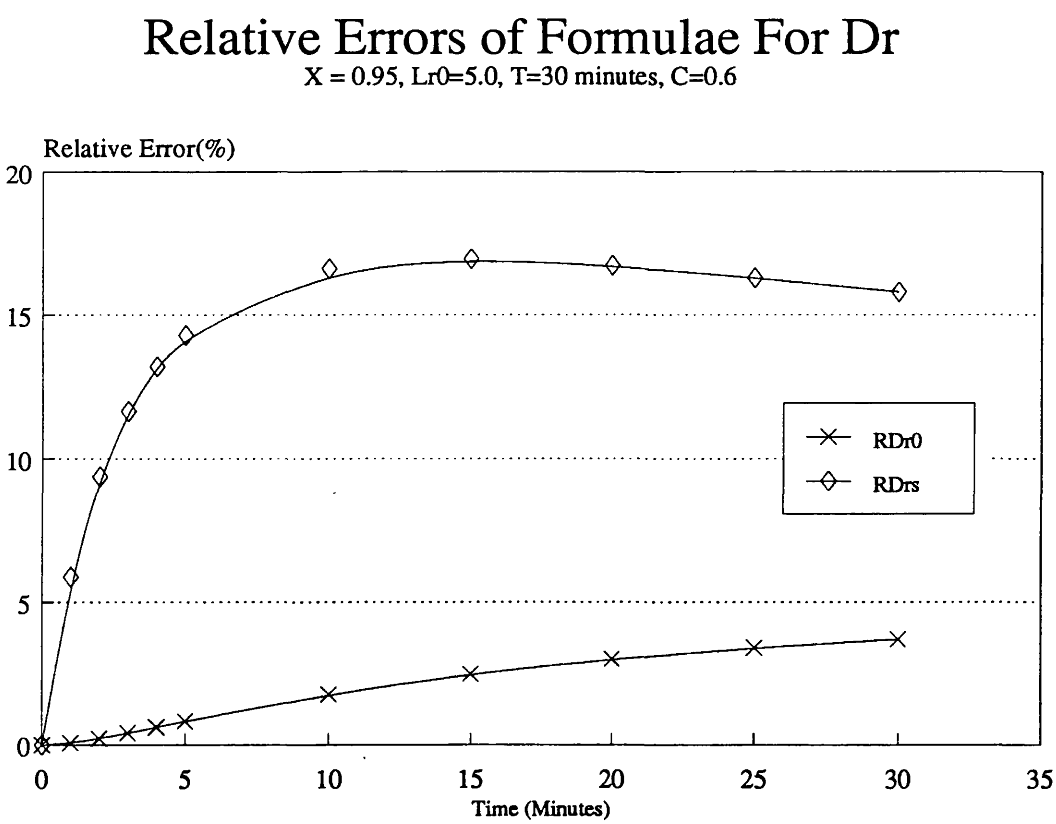


Fig 3.7b

Comparison of Formulae For Dr

X = 0.70, Lr0=5.0, T=30 minutes, C=0.6

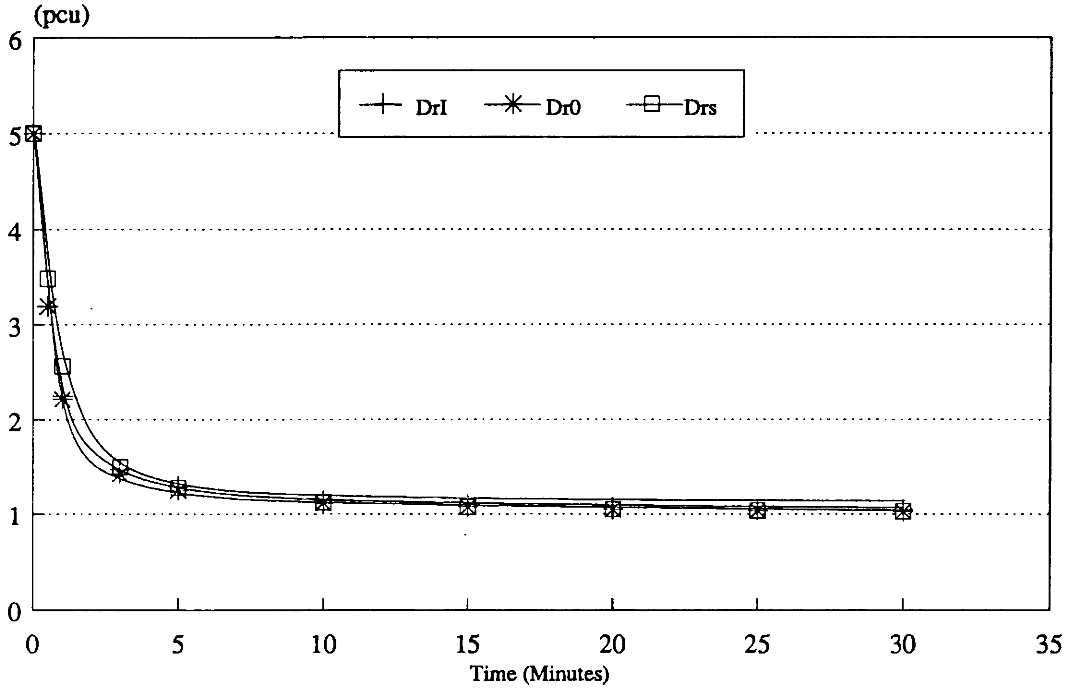


Fig 3.8a

Relative Errors of Formulae For Dr

X = 0.70, Lr0=5.0, T=30 minutes, C=0.6

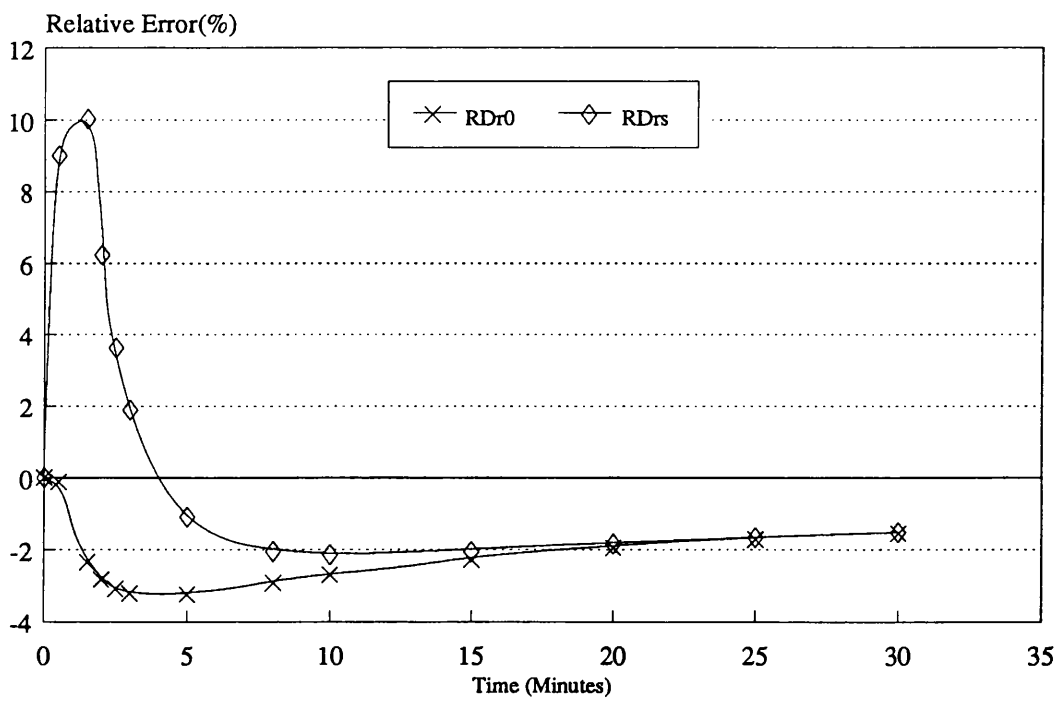


Fig 3.8b

3.4 CONCLUSION

Although the formulae for estimating queues and delays used in OSCADY are not preferable in terms of accuracy, they will still be used in this thesis, since OSCADY is well known, and it it easier to use the same formulae as those used in OSCADY to compare the results given by OSCADY and other methods, when queue length and delay calculations are needed. For the same reason, the parameter C will be taken as 0.6 in this thesis as used in OSCADY.

The formulae used in OSCADY can be summarised as follows, which will be used in this thesis for estimating queues and delays when signal settings are known:

(1) To **estimate** queue length:

$$L = L_u + L_r \tag{3.14}$$

Where L_u is the uniform queue length given by (3.10);
L_r is the random queue length given by (3.7).

(2) To **estimate** delay per unit time:

$$D = D_u + D_r \tag{3.15}$$

Where D_u is the uniform delay per unit time given by (3.10);
D_r is the random delay per unit time given by (3.12).

It should be pointed out that, although some formulae give more accuracy, they cannot be used in the optimatisation process when the gradients with respect to signal timings variables (e.g. Λ) need to be calculated,, since usually these formulae are either non-differentiable or difficult to differentiate. For example, $L_r(t)$ given by (3.7) can be used for estimating the random queue length when the signal settings are known, but when derivatives such as $\partial L_r/\partial\Lambda$ are needed, only (3.5a) can be used. On the other hand, although formulae such as (3.5a) are easy to differentiate, they are inferior to those like (3.7) with respect to accuracy. Hence that in the process of designing signalised road junctions, the formulae for estimation and those for optimisation may be different, although they might be expected to be the same.

CHAPTER 4. TRAFFIC SIGNAL SETTINGS FOR A SINGLE PERIOD

§4.1 INTRODUCTION

The calculation of signal timings, i.e. the cycle time and green times which yield satisfactory operating conditions, is based on the selection of a performance index to represent the effect of the signal settings on the traffic conditions. To calculate traffic signal settings at a junction for a single period in which the arrival rate in each traffic stream is given, the most widely accepted criterion is that of minimising the total delay to the traffic in all streams at the junction.

In this chapter, the difficulties in optimisation when using the delay formulae introduced in Chapter 3 are first identified. The approximate method, as used in the program TRAFSIG, which avoids those difficulties is then introduced. A new approach, based on the idea of coordinate transformation, gives a new extended sheared delay formula which leads to an objective function that is differentiable throughout the range of feasible timings and can be used to calculate the derivatives of the objective function. A stage-based approach is used to optimising the traffic signal settings for a single junction, applying **Allsop's** minimisation routine OPTIM in SIGSET, adapted to the new form of the objective function. Finally, some examples are discussed.

Mathematically, using a stage-based approach as described by **Allsop** (1971), the problem can be stated as follows:

Choose a vector $\underline{\lambda}$ to minimise the objective function

$$W(\underline{\lambda})=T\sum_{j=1}^{n} D_j(\Lambda_j) \tag{4.1}$$

Where n is the number of streams in the junction.

D_j is the delay per unit time for stream j. $(j=1,2,\cdots,n)$

$\underline{\lambda} = (\lambda_0, \lambda_1,\cdots, \lambda_n)$

$$D_j(\Lambda) = D_{uj}(\Lambda) + D_{rj}(\Lambda) \tag{4.2}$$

Equation (4.2) is the same as (3.15) except that the subscript j is introduced.

The effective green time for stream j can be expressed as:

$$\Lambda_j = \sum_{i=0}^{m} a_{ij}\lambda_i \qquad (j=1,2,\cdots,n)$$

Where m is the number of stages in the signal cycle.

And $0 \leq a_{ij} \leq 1 \qquad (j=1,2,\cdots,n)$

$$0 \leq \lambda_i \leq 1 \qquad (i=1,2,\cdots,m)$$

$$0 \leq \Lambda_j \leq 1 \qquad (j=1,2,\cdots,n)$$

$$\lambda_0 = L/c$$

where L is the total lost time for the junction

c is the cycle time.

Subject to constraints (2.23) – (2.28) as introduced in Chapter 2.

The vector $\underline{\lambda}$ will completely determine the signal settings, and the $\underline{\lambda}$ that minimises $D(\underline{\lambda})$ is the optimal solution.

§4.2 DIFFICULTIES IN OPTIMISATION

After the mathematical formulation, $D_j(\Lambda_j)$ should be specified so as to evaluate the gradients of the objective function for use in the optimisation process. In the earlier methods of setting traffic signals, e.g. **Webster's** method and the SIGSET method, only undersaturated junctions are modelled, where equilibrium expressions such as D_{ue} and D_{re} can be used for estimating delay to the traffic.

In order to model the traffic conditions for the whole range of degrees of saturation and, as mentioned in Chapter 3, to be consistent with the program OSCADY, formulae (3.10) and (3.12) should be used to estimate D_{uj}

and D_{rj} in (4.2) when the signal settings are known, and it would be ideal to use those formulae in the optimisation process as well. However, since there is more than one expression for uniform delay D_{uj} when the degree of saturation X_j and the initial queue length L_{r0j} are within different ranges, it would be too complicated to apply (3.10) in the optimisation process when derivatives of the objective function $D_j(\Lambda_j)$ with respect to Λ_j need to be evaluated, although D_{rj} given by (3.12) does not cause any problem in optimisation. So far no optimisation technique is available to solve the problem, except a tedious grid search, which is not practical in terms of computing time.

To optimise signal settings by analytical methods that require the calculation of the derivatives, formulae such as (2.7b) and (2.10) for estimating the uniform delay, which are simpler than (3.10) and widely accepted, are used in the expressions of uniform delay and random delay for optimisation purposes. After optimisation is completed, more accurate expressions can then be employed to evaluate the consequences of the signal settings. Hence D_{uj} can be expressed by the commonly accepted formulae (3.8) and (3.9) introduced in §3.2.2:

$$D_{uj}(\Lambda_j) = \begin{cases} D_{uej}(\Lambda_j) = \dfrac{X_j Q_j c(1-\Lambda_j)^2}{2(1-\Lambda_j X_j)} & \text{When } 0 \le X_j < 1 \\ D_{uoj}(\Lambda_j) = \dfrac{Q_j c(1-\Lambda_j)}{2} & \text{When } X_j \ge 1 \end{cases} \tag{4.3}$$

However, a problem still exists after the simplification, since $D_{uj}(\Lambda_j)$ is non-convex and is non-differentiable with respect to Λ_j at $\Lambda_j = y_j$ (i.e. $X_j = 1$), $D_j(\Lambda_j)$ has the same property (Fig 4.1). This leads to problems in optimisation when gradient methods are employed. Methods are therefore needed to overcome this difficulty.

§4.3 AVOIDING THE DIFFICULTIES USING EXISTING DELAY EXPRESSIONS

As stated in §2.4.2, one solution to solve the problem outlined above is to use D_{uej} as the expression for estimating uniform delay for the whole range of degrees of saturation. i.e. let $D_{uj}(\Lambda_j) = D_{uej}(\Lambda_j)$ for any X_j. This is obviously an overestimation of $D_{uj}(\Lambda_j)$ when $\Lambda_j < y_j$, but since $D_{uj}(\Lambda_j)$ is convex and differentiable now, gradient methods can come into force. The program TRAFSIG uses this approach when total delay is used as the

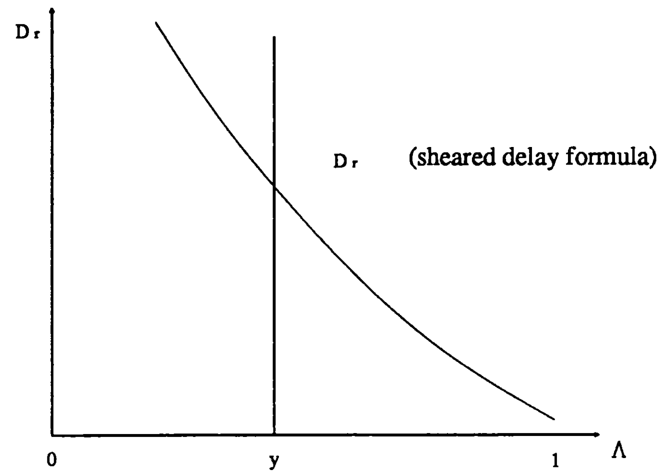


Fig 4.1a The Random Delay

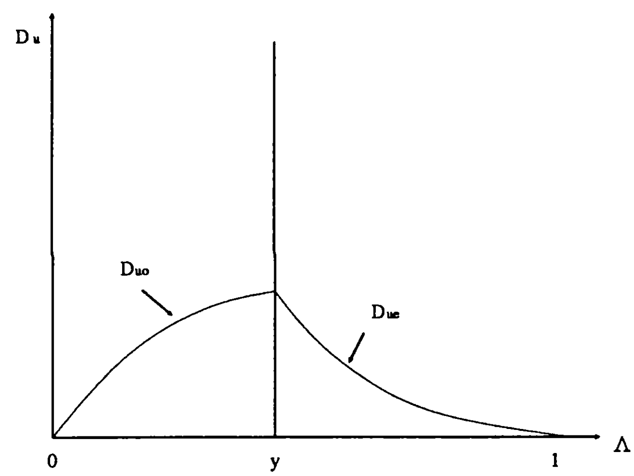


Fig 4.1b The Uniform Delay

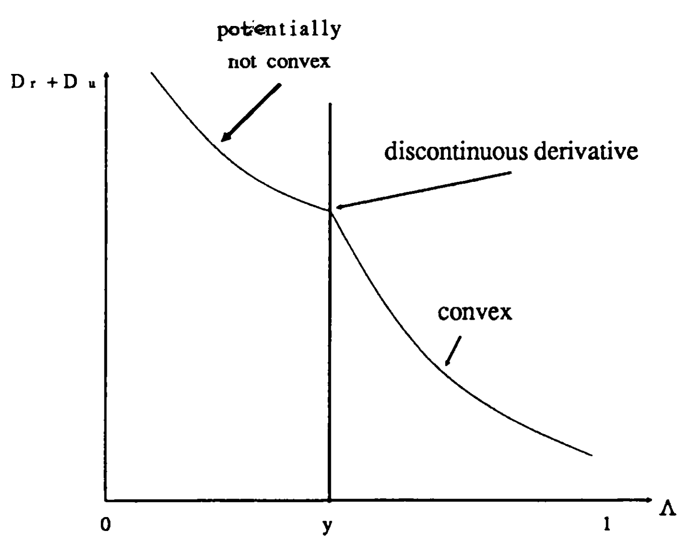


Fig 4.1c The (Random + Uniform) Delay

objective function. However, this method may bring errors into the signal settings obtained, especially when $D_{uoj}(\Lambda_j)$ and $D_{uej}(\Lambda_j)$ differ substantially. However, it gives an approximation to the signal timings that a more accurate method would produce, and the difference between $D_{uoj}(\Lambda_j)$ and $D_{uej}(\Lambda_j)$ though substantial compared with $D_{uoj}(\Lambda_j)$, will often be small compared with $D_{rj}(\Lambda_j)$ when $\Lambda_j<y_j$. The validity of this method will be discussed in the examples at the end of this chapter.

§4.4 A NEW APPROACH

§4.4.1 Introduction

In Table 3.1 it was shown that so far no comprehensive formula has been available for estimating D_u or (D_u+D_r). However, the difficulty of non-differentiablity can be overcome if we can find a comprehensive formula for estimating (D_u+D_r).

In this thesis we aim to develop a comprehensive delay formula for estimating (D_u+D_r), so that the optimisation can be implemented and the vacancy in Table 3.1 can be filled. The approach is based on the coordinate transformation technique developed by **Kimber** and **Hollis** (1979).

§4.4.2. New Expression For Delay

Enlightened by **Kimber** & **Hollis**'s approach of shearing the random delay formula, we try to establish a comprehensive delay formula, taking into account both the random delay and the uniform delay.

In the following part, for each stream j, the subscript j is omitted from the relevant notation and the argument (Λ_j) from the various delay functions. For example, Λ stands for Λ_j, D_{ue} for $D_{uej}(\Lambda_j)$, etc.

Before introducing the method, further to Table 3.1, we first summarise the formulae available for estimating the delays by steady-state theory and deterministic theory in Table 4.1, where those formulae are explicitly given.

D_r and D_u are plotted against X with Λ regarded as fixed and q as variable in Fig 4.2a and Fig 4.2b respectively, with arrivals regarded as uniform when $X \geq 1$ in Fig 4.2a, so that D_{ro} represents only the overload delay, without allowance for randomness of arrivals. Then in Fig 4.2c, when $X = X_o \geq 1$, the arrivals are regarded as uniform so that the delay D can be expressed as:

$$D = D_o = D_{ro} + D_{uo} = L_{r0} + 0.5QT(X_o - 1) + \frac{Qc(1-\Lambda)}{2} \tag{4.4}$$

Table 4.1 Expressions For Estimating Delays

		Equilibrium ($X < 1$)	**Overload** ($X \geq 1$)
DELAY D	**Uniform** D_u	$D_{ue} = \dfrac{qc(1-\Lambda)^2}{2(1-y)}$	$D_{uo} = \dfrac{Qc(1-\Lambda)}{2}$
	Random D_r	$D_{re} = \dfrac{CX_e^2}{1-X_e}$	$D_{ro} = L_{r0} + 0.5QT(X_o - 1)$ if arrivals are regarded as uniform

Hence we have

$$X_o = \frac{2D_o - 2L_{r0} - Qc(1-\Lambda)}{QT} + 1 \tag{4.5}$$

On the other hand, when $X = X_e < 1$ and the randomness of arrivals is allowed for, the delay D can be expressed as:

$$D = D_e = D_{re} + D_{ue} = \frac{CX_e^2}{1 - X_e} + \frac{qc(1-\Lambda)^2}{2(1-y)} \tag{4.6}$$

Therefore D can be expressed differently when $X \geq 1$ and $X < 1$, and according as the randomness in arrivals is or is not allowed for.

The coordinate transformation is then illustrated by Fig 4.2d. Let $AA' = BB'$ as a typical line of constant D, where B is the intersection with the desired sheared curve expressing D as a function of the abscissa X of B, i.e. let

$$1 - X_e = X_o - X \tag{4.7a}$$

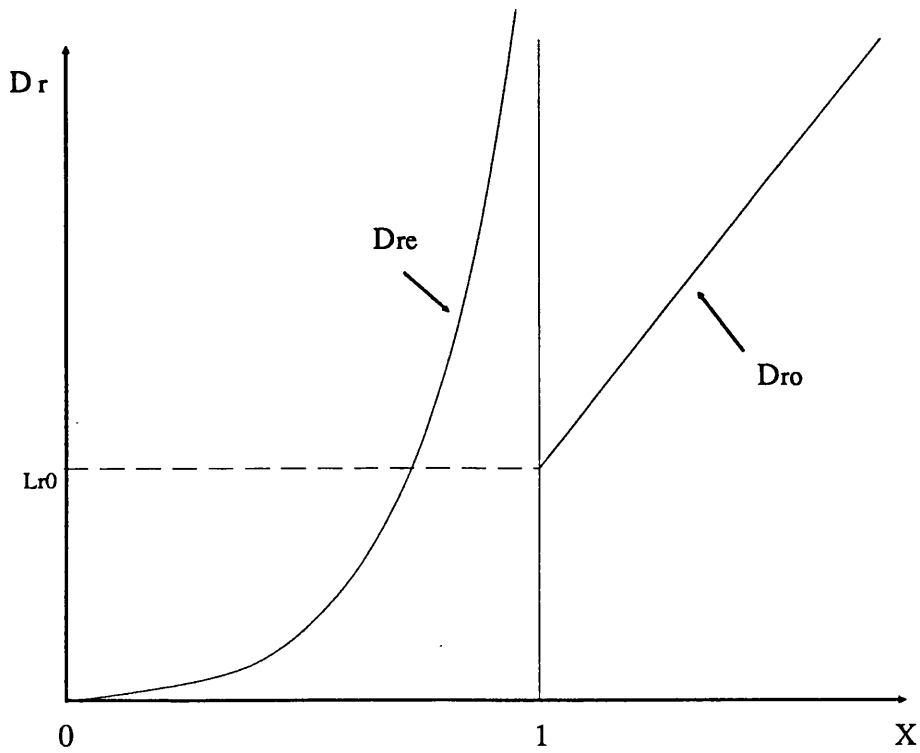


Fig 4.2a The Random Delay

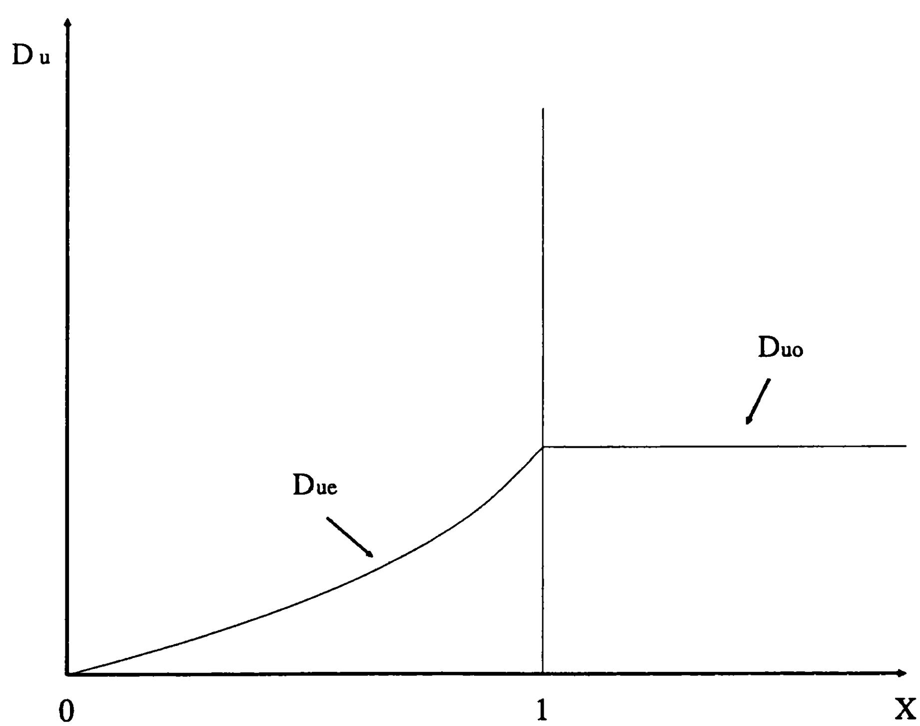


Fig 4.2b The Uniform Delay

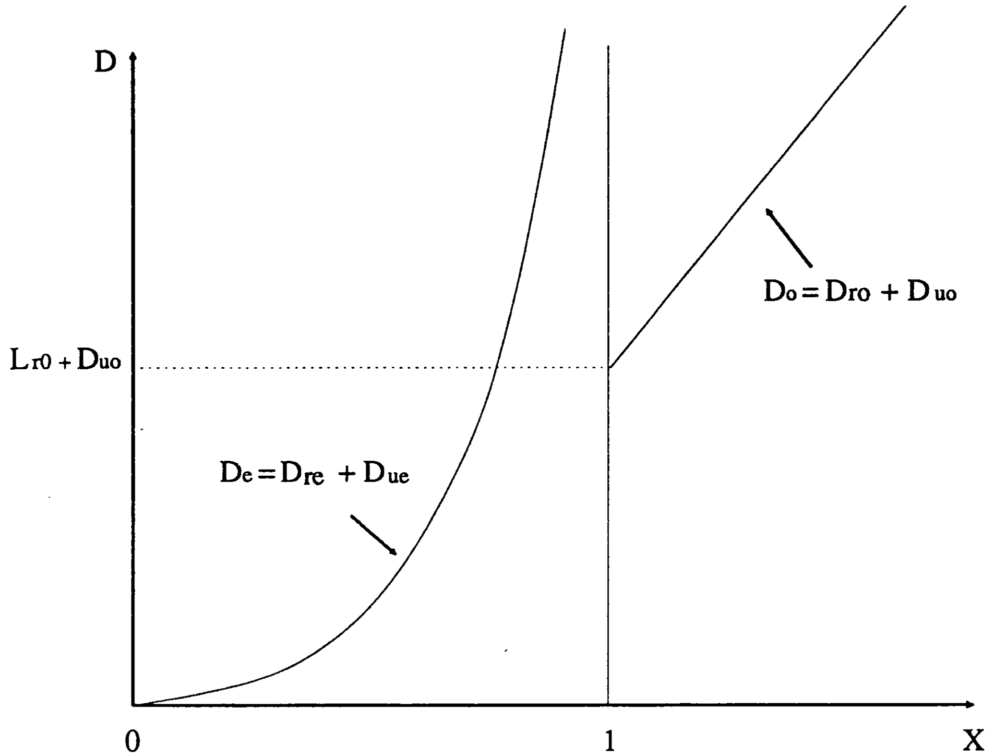


Fig 4.2c The Random Delay + Uniform Delay
Over A Time Period T

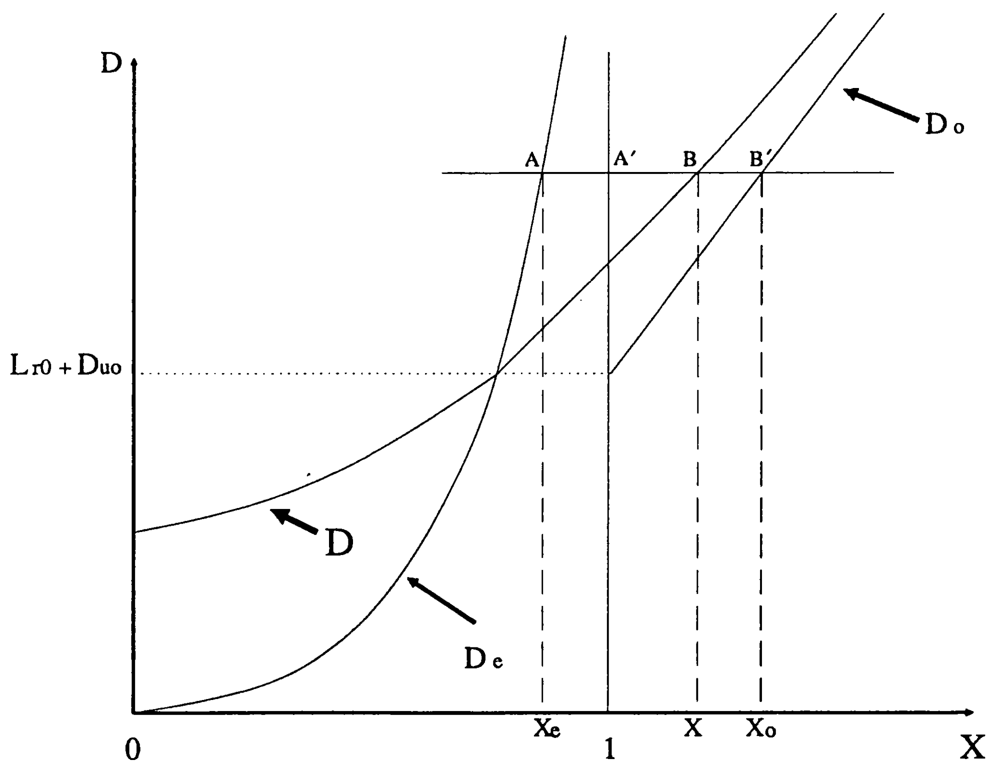


Fig 4.2d The Development Of The New Delay Formula

This relationship is used together with (4.5) to express D in terms of X_e as follows:

$$X_e = 1 - X_o + X = A - \frac{2D}{QT} \tag{4.7b}$$

where
$$A = \frac{qT + 2L_{r0} + Qc(1-\Lambda)}{QT} \tag{4.8}$$

and
$$q = XQ$$

Hence $D = 0.5QT(A - X_e)$ (4.9)

Substituting this expression for D, $y=\Lambda X_e$ and $q = X_e Q$ into (4.6), we get a cubic equation:

$$a_0 X_e^3 + a_1 X_e^2 + a_2 X_e + a_3 = 0 \tag{4.10a}$$

where:

$$a_0 = QT\Lambda - 2C\Lambda \tag{4.11a}$$

$$a_1 = 2C - Qc(1-\Lambda)^2 - QT(\Lambda A + \Lambda + 1) \tag{4.11b}$$

$$a_2 = Qc(1-\Lambda)^2 + QT(\Lambda A + A + 1) \tag{4.11c}$$

$$a_3 = -QTA \tag{4.11d}$$

According to the theory of equations (see eg **Lovitt** 1939, Appendix 1), the solution for X_e can be discussed as follows:

1. When $T \neq 2C/Q$, then $a_0 \neq 0$, and equation (4.10) becomes:

$$X_e^3 + b_1 X_e^2 + b_2 X_e + b_3 = 0 \tag{4.12a}$$

where $b_i = a_i/a_0$, $i=1,2,3$ (4.12b)

Let $X_e = Z - b_1/3$ (4.13)

we have

$$Z^3+p_1Z+p_2=0 \tag{4.14a}$$

where $p_1=b_2-b_1^2/3$, (4.14b)

$$p_2=b_3-b_1b_2/3+2b_1^3/27 \tag{4.14c}$$

The discriminate $\Delta=p_1^3/27+p_2^2/4$ was investigated. Attempts to show algebraically that it is less than zero for relevant values were unsuccessful, but it was found to be negative when tested by numerical evaluation for a wide range of curves with $C=0.6$, $L_{r0}\geq0$, $500\leq s\leq7000$, $X>0$, $0<c<120$, $0<\Lambda<1$ and $T>0$. Hence it appears that there are three distinct real roots of equation (4.14a) in Z, and so is the case for the equation (4.12a) in X_e.

Let $$U=-\frac{p_2}{2}+i\sqrt{-\Delta} \tag{4.15a}$$

$$V=-\frac{p_2}{2}-i\sqrt{-\Delta} \tag{4.15b}$$

U can also be expressed as:

$$U = R^{i\theta} \tag{4.15c}$$

Where $$R=\sqrt{(-p_2/2)^2+\{\sqrt{-\Delta}\}^2}$$

$$\theta=\mathrm{ArcCos}\left\{\frac{-p_2}{2R}\right\}$$

Then let $U^{1/3}=U_1 + iU_2$ (4.16a)

Where

$$U_1=\sqrt[3]{R}\ \mathrm{Cos}(\theta/3) \tag{4.16b}$$

$$U_2=\sqrt[3]{R}\ \mathrm{Sin}(\theta/3) \tag{4.16c}$$

The three roots for Z are:

$$Z_2=-U_1-U_2\sqrt{3}= -2\sqrt[3]{R}\ \mathrm{Sin}(\theta/3 + \pi/6) \qquad (4.17a)$$

$$Z_2=-U_1-U_2\sqrt{3}= -2\sqrt[3]{R}\ \mathrm{Sin}(\theta/3 + \pi/6) \qquad (4.17b)$$

$$Z_3=-U_1+U_2\sqrt{3}= 2\sqrt[3]{R}\ \mathrm{Sin}(\theta/3 - \pi/6) \qquad (4.17c)$$

Since $0<\theta<\pi$, we have $0<\theta/3<\pi/3$, hence $U_2/U_1=tg(\theta/3)<tg(\pi/3)=\sqrt{3}$, i.e. $U_2<U_1\sqrt{3}$, hence $2U_1>-U_1+U_2\sqrt{3}$, therefore $Z_2<Z_3<Z_1$, i.e. Z_2 is the smallest root, Z_3 is the middle root, and Z_1 is the biggest root.

By calculating the corresponding X_e it was found that:

(i) When $T > 2C/Q$, only Z_2 gives a value of X_e satisfying the condition: $0 < X_e < 1$. Hence when $T > 2C/Q$, we take $Z = Z_2$ as the solution for Z. This is the usual case since 2C/Q is very small and normally T is within the range 10~15 minutes, which far exceeds 2C/Q.

(ii)When $T < 2C/Q$, only Z_3 gives a value of X_e satisfying the condition $0 < X_e < 1$. Hence when $T < 2C/Q$, we take $Z = Z_3$ as the solution for Z. This solution is also useful when delay needs to be evaluated for very small time intervals, as in the case of making a time-shift to be introduced in the next chapter.

Thus when $T \neq 2C/Q$, Z can be given either by Z_2 or by Z_3, and the corresponding X_e can then be obtained from the relationship (4.13a).

2. However, When $T = 2C/Q$, then $a_0=0$, therefore (4.10a) becomes a quadratic equation:

$$a_1X_e^2 + a_2X_e + a_3 = 0 \qquad (4.18)$$

And the solution for X_e will simply be:

$$X_e = \frac{-a_2+\sqrt{a_2^2-4a_1a_3}}{2a_1} \qquad (4.19)$$

Some of these numerical results can also be proved algebraically. Define $f(X) = a_0X^3+a_1X^2+a_2X+a_3$, then

$$f'(X) = 3a_0X^2+2a_1X+a_2$$

$$f''(X) = 2(3a_0X+a_1)$$

And since $f(0)=a_3<0$, $f(1)=a_0+a_1+a_2+a_3=2C(1-\Lambda)>0$, hence there is either only one real root for X in (0,1) or there are three.

When $a_0=0$ (i.e. QT=2C), $a_1<0$, and $f(X) \to -\infty$ as $X \to \pm\infty$, but $f(1)>0$, hence $f(X)$ must cross the line $f(X)=0$ at least once in $(1, \infty)$. Since there are only two real roots in this case, there is only one real root in (0, 1), and this should be the smaller root.

When $a_0<0$ (i.e. QT<2C), $f(X)\to\infty$ as $X\to-\infty$, and $f(X)\to-\infty$ as $X\to\infty$, but $f(0)<0$ and $f(1)>0$, hence $f(X)$ must cross the line $f(X)=0$ at least once each in the regions $(-\infty, 0)$ and $(1, \infty)$, hence there is only one real root in (0,1), and this is the middle root corresponding to Z_3.

When $a_0>0$ (i.e. QT>2C), $f(X)\to-\infty$ as $X\to-\infty$, and $f(X)\to\infty$ as $X\to\infty$, $f'(X)\to\infty$ as $X\to-\infty$, and $f'(X)\to\infty$ as $X\to\infty$, $f'(0)=a_2>0$, $f''(0)=2a_1<0$ and as before $f(0)<0$ and $f(1)>0$.

1) If $f''(1)=2\cdot(3a_0+a_1) \leq 0$, then $3a_0X+a_1<0$ for all $0<X<1$, i.e. $f''(X)<0$ in (0,1), hence $f(X)$ is concave in (0,1). Therefore there is only one root in (0,1) and this is the smallest root corresponding to Z_2.

2) If $f''(1)=2\cdot(3a_0+a_1) >0$, then consider the zeros of $f'(X)$. These are:

$$0< X_1=\frac{-a_1-\sqrt{a_1^2-3a_0a_2}}{3a_0} < X_2=\frac{-a_1+\sqrt{a_1^2-3a_0a_2}}{3a_0}$$

i) If $f'(1)=3a_0+2a_1+a_2<0$, then $X_1<1$ and $X_2>1$, and the largest two roots of $f(X) >1$, hence there is only one root in (0,1), and this is the smallest root corresponding to Z_2.

ii) If $f'(1)=3a_0+2a_1+a_2 \geq 0$, then there may be three real roots in (0,1), and by continuity the relevant one must be the one corresponding to Z_2.

Hence X_e can be obtained whatever the difference (T-2C/Q) is. Then by the relationship

$$D = 0.5QT(A-X_e) \qquad (4.20)$$

We have the final expression for D in terms of L_{r0}, Q, T, Λ and q. This expression is derived from a diagram in which D is plotted against X =q/Q for fixed Λ and hence fixed Q with q regarded as variable. It will, however, be used as an expression for D in terms of Λ for fixed q and with $Q = \Lambda s$.

§4.4.3 Acceptability And Convexity

The acceptability of the extended sheared delay formula as an approximation to delay can be evaluated by calculating the difference between this formula and the formula which is used in OSCADY for estimating (but not for minimising) delay. Extensive calculations for different values of the parameters have shown that the differences are usually very small, especially when Λ is near to y, where the optimum is often located. An example of comparing the two formulae is given in Fig 4.3, where the data for the stream is: q=900 pcu/hour, s=2000 pcu/hour, L_{r0}=0.0 pcu, and T=5 minutes. The notation is:

D_y = OSCADY formula, where the uniform component D_{yu} is given by (3.10), and the random component D_{yr} is given by **Kimber & Hollis**'s (1979) sheared delay formula (3.12);

D_n = The new sheared delay formula (4.20); and

D_s = The formula whose uniform component D_{su} is given by the simplified uniform delay expression (4.3), and the random component $D_{sr}=D_{yr}$.

D_s is plotted here together with D_y and D_n since it is closely related to D_n. It can be seen from Fig 4.3 that the absolute difference between those formulae is rather small. The extensive calculations show that in most cases the absolute differences are less than 1 pcu, although the relative

Comparison of Delay Formulae
q=900 pcu/h, s=2000pcu/h, c=120s, T=5min
Lr0=0

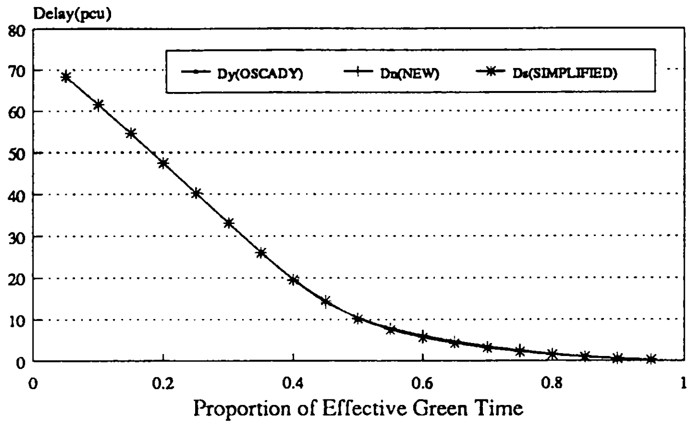


Fig 4.3a

Comparison of Delay Formulae
q=900 pcu/h, s=2000pcu/h, c=120s, T=5min
Lr0=0

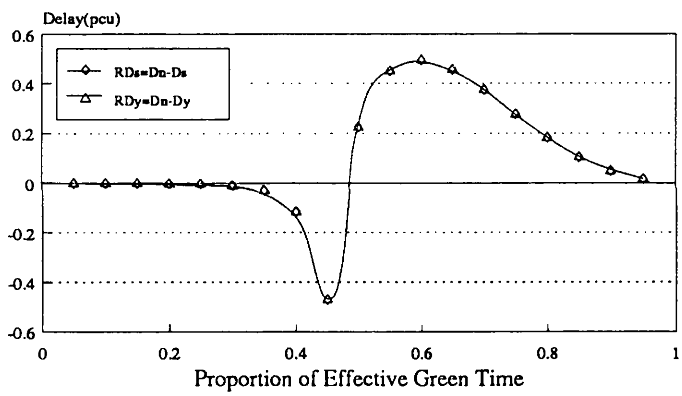


Fig 4.3b Absolute Differences Between
Dn and Dy, Ds

Comparison of Delay Formulae
q=900 pcu/h, s=2000pcu/h, c=120s, T=5min
Lr0=0

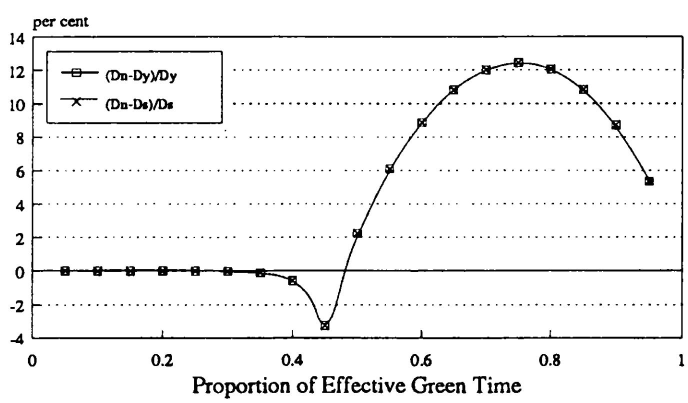


Fig 4.3c Relative Differences Between
Dn and Dy, Ds

difference can be big when the delay is very small. This means that the extended sheared delay formula can be used for optimisation when gradient methods are necessary, so that the difficulty of discontinues derivatives can be overcome.

Although the extended sheared delay formula is not strictly convex for all the ranges of Λ and $\lambda_0 = 1/c$ (where c is the cycle time), it is nearly convex in most cases and is usually convex when Λ is around y. Extensive calculations for D_n for a wide range of T, s, Y and c were made and the results show that in most cases D_n is well behaved (Fig 4.4a), and even if D_n is non-convex, the non-convexity only occurs when Λ is much less than y, i.e. D_n is non-convex only when X is very big (Fig 4.4b). In rare cases D_n is just slightly non-convex when $\Lambda > y$ (Fig 4.4c). This will rarely lead to problems because the objective function is the sum of the rates of delay for a number of streams and not all the delay curves are likely to be non-convex in the same region, so the likelihood that a non-convex surface will result is small.

§4.5 THE OPTIMISATION METHOD

The optimisation technique is based on the optimisation subroutine OPTIM that was used in SIGSET (Allsop 1971) and is modified for the current objective function. The derivatives of the objective function with respect to Λ and λ_0 are given in Appendix 2. Derivatives with respect to the λ_i $(1 \leq i \leq m)$ follow by the chain rule because each Λ_j is a known linear combination of the λ_i (2.21). OPTIM is used here because all the variables, constraints and the framework of signal timing requirements are the same as in SIGSET, except the change in the objective functions, where in SIGSET Webster's two term delay expression (2.2) is used, and the objective function is convex; here the new time-dependent delay expression (4.20) is used and the objective function may be non-convex but is likely to be convex in most cases. The example calculations introduced below in §4.6 suggest that the possibility of non-convexity will not cause problems in practice and the optimiser can be used for optimising traffic signal settings for a single time period.

Convexity of Delay Formulae

y=0.5, s=2960 pcu/h, c=120s, T=5 min

Lr0=0

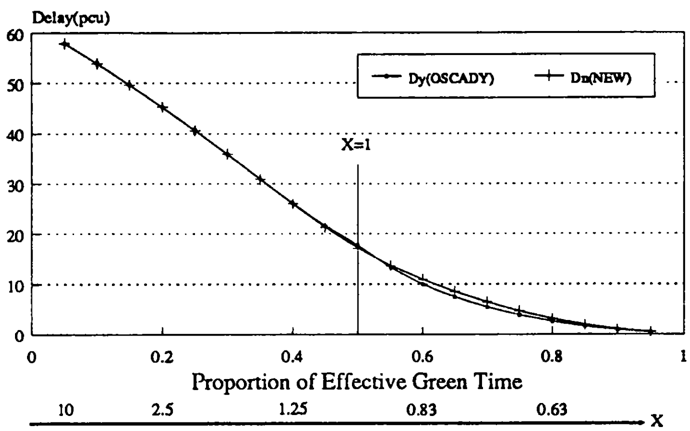


Fig 4.4a

Convexity of Delay Formulae

y=0.8, s=2960 pcu/h, c=120s, T=5 min

Lr0=0

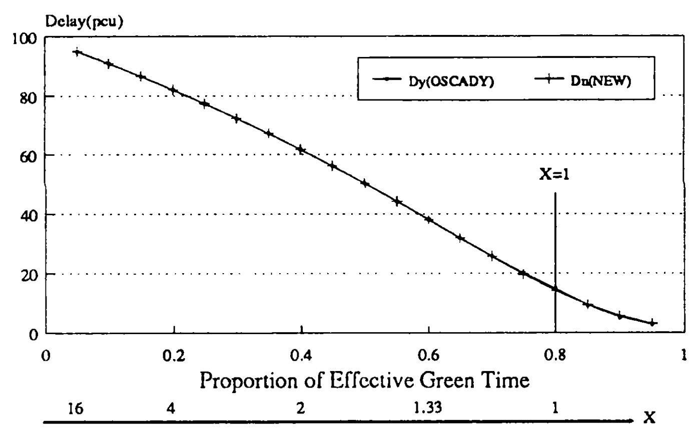
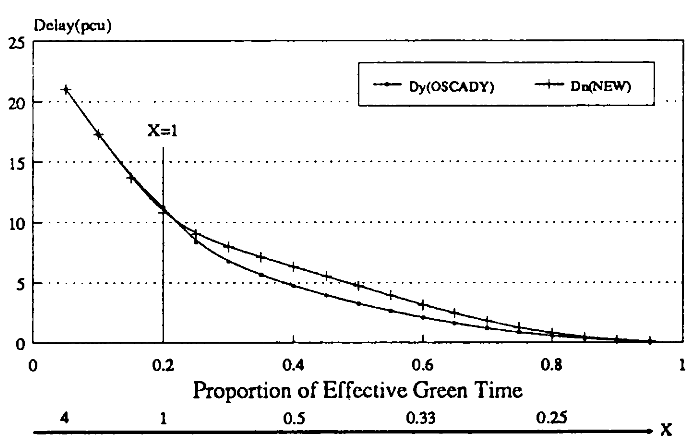


Fig 4.4b

Convexity of Delay Formulae

y=0.2, s=2960 pcu/h, c=120s, T=5 min

Lr0=0



F4.4c

§4.6 EXAMPLES

§4.6.1 Introduction

The aim of developing a new delay expression is to overcome the difficulties in optimising traffic signals when considering the full range of the degree of saturation rather than merely the unsaturated situations. As mentioned earlier, the existing method SIGSET, although it can optimise signal settings for undersaturated junctions, cannot be used for oversaturated cases due to the formula used for estimating delay. The method that avoids the difficulties by using existing delay expression, can only give approximate solutions, therefore the signal settings resulting from it should be evaluated. OSCADY calculates settings on the basis of capacity maximisation when the junction is overloaded by the given arrival rates.

In order to illustrate the applicability of the present method, calculations of signal settings are made for three example junctions using:

(1) The NEW method using the new delay expression;
(2) The SIGSET method (when the junction is undersaturated);
(3) The OSCADY method; and
(4) The SIMPLE method that avoids the optimisation difficulties in the manners described in §4.3.

Except in OSCADY when the junction is overloaded, the subroutine OPTIM in SIGSET is used in all the methods to optimise the signal settings. However, the calculations of the derivatives of the objective function (4.1) with respect to λ_i $(i=1,2,\cdots,m)$ are based on different formulae:

(1) The NEW method: using (4.20);

(2) The SIGSET method
(3) The OSCADY method
} using **Webster**'s two term delay expression (2.2);

(4) The SIMPLE method: using **Kimber** & **Hollis**'s (1979) sheared delay formula (3.12) and the equilibrium uniform delay expression (3.8).

It should be noted that since the OSCADY output gives rounded signal timings, some discrepancies can occur when making the comparison. Nevertheless, a good general idea of differences between results given by the four methods can be obtained.

The first two junctions are simple crossroads, the third one is a complex junction. The cases that each $X_j < 1$, and that some $X_j < 1$ and some $X_j \geq 1$ are considered for each junction, so that the full range of degrees of saturation can be modelled. In every case, a maximum cycle time of 120 seconds is assumed. The length of time period, T, is chosen as 30 minutes for all the examples in every case.

When each stream in the junction is undersaturated, i.e. $X_j<1$, $j=1,2,\cdots,n$, where n is the number of streams, calculations are made for all four methods. However, when at least one of the streams is oversaturated, i.e. there exists at least one j, such that $X_j \geq 1$, $j=1,2,\cdots,n$, then calculations are only made for three methods, since in this case the SIGSET method cannot be applied.

After obtaining the signal timings by each method, in order to compare those methods, total delay, i.e. delay per unit time multiplied by T minutes, can be used as a performance index. Since these methods use different expressions for calculating the derivatives of delay per unit time with respect to $\underline{\lambda}$, there is no common expression to evaluate the delay as a result of the signal settings for each method. In this case the expressions used in OSCADY method are taken to estimate the total delay W and queue lengths. Another important factor in comparing the results is the junction reserve capacity σ. The maximum acceptable degree of saturation for each stream, P_j $(j=1,2,\cdots,n)$ is taken as 0.9. Other auxiliary indications of performance of the junctions are the X_j, the degrees of saturation for each stream, and the L_{fj} $(=L_{rfj}+L_{ufj})$, the final queue length for each stream.

§4.6.2 Junction 1 — Symmetrical Crossroads of Two One-way Streets

This junction is the simplest to start with. It has only two streams, therefore there are only two stages in the signal cycle. Each stream has a saturation flow of 2000 pcu/h as shown in Fig 4.5.

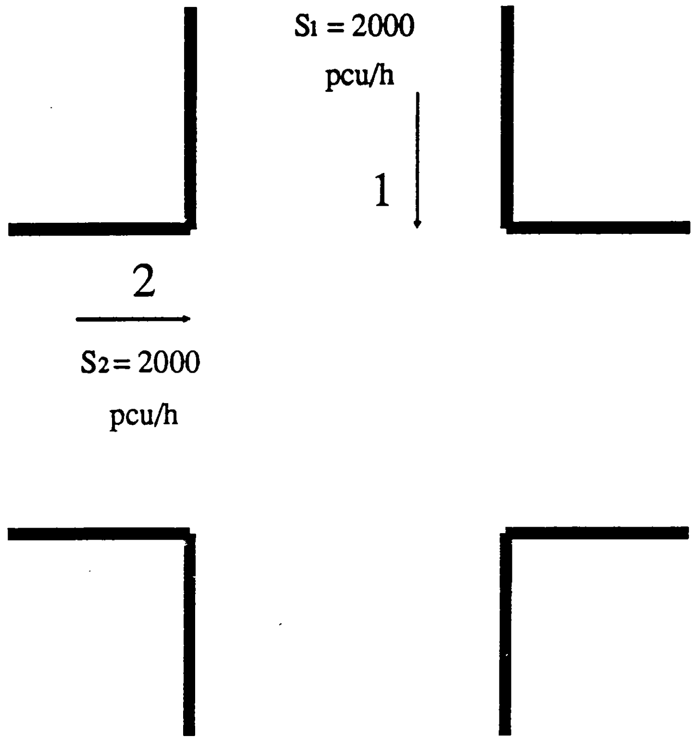


Fig 4.5 Junction 1

Other parameters for this junction are:

Table 4.2

Stage		1	2
Minimum Green	(Seconds)	*6.00*	*6.00*
Lost Time After Stage	(Seconds)	*4.00*	*4.00*

Case 1: each $X_j < 1$

Flow ratios $y_1 = 0.45$, $y_2 = 0.3$

Table 4.3

METHOD		SIGSET	SIMPLE	NEW	OSCADY
Cycle Time (Seconds)		70.35	63.07	64.87	70.40
Proportion of Cycle Effectively Green For Stage	1	0.5268	0.5200	0.5220	0.5256
	2	0.3595	0.3532	0.3547	0.3594
Proportion of Cycle Effectively Green For Stream	1	0.5268	0.5200	0.5220	0.5256
	2	0.3595	0.3532	0.3547	0.3594
Degree of Saturation In Stream (%)	1	85.42	86.54	86.21	85.62
	2	83.46	84.95	84.58	83.48
Total Queue Length In Stream (pcu)	1	6.44	6.45	6.43	6.51
	2	5.83	5.83	5.83	5.83
Delay Per Unit Time In Stream (pcu)	1	6.32	6.30	6.28	6.38
	2	5.71	5.68	5.69	5.72
Total Delay (pcu·mins)		360.8	359.3	359.1	363.0
Reserve Capacity		5.36%	4.00%	4.40%	5.11%

It can be seen from Table 4.3 that the SIGSET method and OSCADY method give longer cycle times and more reserve capacity than the SIMPLE and NEW method, but just slightly higher total delay. The NEW method gives fractionally less delay and a little more reserve capacity compared with the SIMPLE method. The ratio of the green times for the two streams is very similar with all four methods.

Case 2: some $X_j \geq 1$

Flow ratios $y_1 = 0.6$, $y_2 = 0.45$

Table 4.4

METHOD		SIMPLE	NEW	OSCADY
Cycle Time (Seconds)		120.00	120.00	120.00
Proportion of Cycle Effectively Green For Stage	1	0.5659	0.5657	0.5600
	2	0.3674	0.3676	0.3733
Proportion of Cycle Effectively Green For Stream	1	0.5659	0.5657	0.5600
	2	0.3674	0.3676	0.3733
Degree of Saturation In Stream (%)	1	106.02	106.06	107.14
	2	108.87	108.81	107.14
Total Queue Length In Stream (pcu)	1	50.15	50.30	55.19
	2	45.93	45.77	40.98
Delay Per Unit Time In Stream (pcu)	1	32.02	32.09	34.36
	2	28.96	28.89	26.71
Total Delay (pcu·mins)		1829.3	1829.3	1832.0
Reserve Capacity		-17.33%	-17.29%	-16.00%

In this case since the junction is oversaturated, all the three methods applied result in a maximum 120 second cycle time. The SIMPLE and NEW methods perform delay minimisation, and give very similar signal settings, but OSCADY calculates the signal timings by capacity maximisation, so that the junction is somewhat less oversaturated with its settings. The resulting extra delay with the OSCADY settings is, however, marginal.

§4.6.3 Junction 2 — Asymmetrical Crossroads of Two One-way Streets

In this case, one of the streams has a saturation flow of 4000 pcu/h as shown in Fig 4.6.

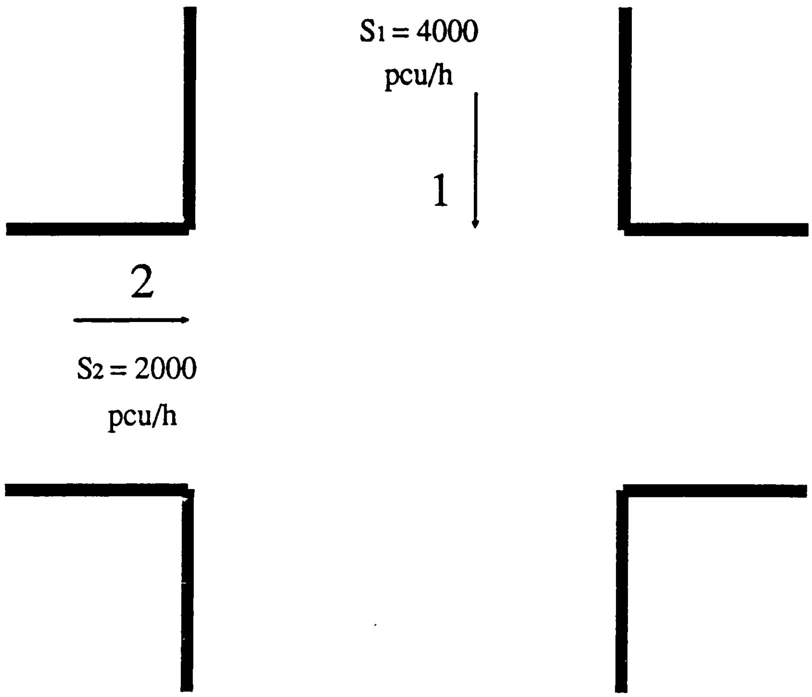


Fig 4.6 Junction 2

Other parameters for this junction are:

Table 4.5

Stage		1	2
Minimum Green	(Seconds)	*6.00*	*6.00*
Lost Time After Stage	(Seconds)	*4.00*	*4.00*

Case 1: each $X_j < 1$

(1) Flow ratios: $y_1=0.3$, $y_2=0.45$

Table 4.6

METHOD		SIGSET	SIMPLE	NEW	OSCADY
Cycle Time (Seconds)		63.29	58.38	59.72	63.50
Proportion of Cycle Effectively Green For Stage	1	0.3566	0.3530	0.3545	0.3567
	2	0.5170	0.5100	0.5115	0.5173
Proportion of Cycle Effectively Green For Stream	1	0.3566	0.3530	0.3545	0.3567
	2	0.5170	0.5100	0.5115	0.5173
Degree of Saturation In Stream (%)	1	84.12	84.98	84.62	84.11
	2	87.05	88.24	87.98	86.99
Total Queue Length In Stream (pcu)	1	8.83	8.60	8.62	8.85
	2	6.65	6.87	6.83	6.64
Delay Per Unit Time In Stream (pcu)	1	8.75	8.52	8.54	8.77
	2	6.48	6.64	6.62	6.47
Total Delay (pcu·mins)		457.0	454.7	454.7	457.2
Reserve Capacity		3.39%	1.99%	2.30%	3.46%

(2) Flow ratios y_1=0.45, y_2=0.3

Table 4.7

METHOD		SIGSET	SIMPLE	NEW	OSCADY
Cycle Time (Seconds)		64.15	59.03	60.37	64.00
Proportion of Cycle Effectively Green For Stage	1	0.5234	0.5203	0.5220	0.5234
	2	0.3519	0.3441	0.3455	0.3516
Proportion of Cycle Effectively Green For Stream	1	0.5234	0.5203	0.5220	0.5234
	2	0.3519	0.3441	0.3455	0.3516
Degree of Saturation In Stream (%)	1	85.98	86.48	86.21	85.97
	2	85.24	87.17	86.84	85.33
Total Queue Length In Stream (pcu)	1	9.70	9.40	9.41	9.68
	2	5.96	6.26	6.23	5.98
Delay Per Unit Time In Stream (pcu)	1	9.62	9.31	9.33	9.60
	2	5.80	6.03	6.01	5.82
Total Delay (pcu·mins)		462.8	460.1	460.0	462.6
Reserve Capacity		4.67%	3.24%	3.64%	4.69%

In both combinations of the flow ratio y_1 and y_2, again the SIMPLE and NEW methods give shorter cycle times, and less reserve capacity compared with the SIGSET and OSCADY methods. The NEW method gives slightly longer cycle time and more reserve capacity than the SIMPLE method, and each gives just slightly lower total delay than SIGSET and OSCADY.

Case 2: some $X_j \geq 1$

(1) Flow ratios y_1= 0.4, y_2= 0.6

Table 4.8

METHOD		SIMPLE	NEW	OSCADY
Cycle Time (Seconds)		120.00	120.00	120.00
Proportion of Cycle Effectively Green For Stage	1	0.4039	0.4043	0.3733
	2	0.5294	0.5290	0.5600
Proportion of Cycle Effectively Green For Stream	1	0.4039	0.4043	0.3733
	2	0.5294	0.5290	0.5600
Degree of Saturation In Stream (%)	1	99.04	98.92	107.14
	2	113.33	113.43	107.14
Total Queue Length In Stream (pcu)	1	33.57	33.17	76.18
	2	83.05	83.48	55.19
Delay Per Unit Time In Stream (pcu)	1	28.89	28.66	48.78
	2	47.52	47.72	34.36
Total Delay (pcu·mins)		2292.0	2291.6	2494.2
Reserve Capacity		-20.58%	-20.65%	-16.00%

(2) Flow ratios $y_1 = 0.6$, $y_2 = 0.4$

Table 4.9

METHOD		SIMPLE	NEW	OSCADY
Cycle Time (Seconds)		120.00	120.00	120.00
Proportion of Cycle Effectively Green For Stage	1	0.6069	0.6074	0.5600
	2	0.3264	0.3259	0.3733
Proportion of Cycle Effectively Green For Stream	1	0.6069	0.6074	0.5600
	2	0.3264	0.3259	0.3733
Degree of Saturation In Stream (%)	1	98.86	98.78	107.14
	2	122.54	122.73	107.14
Total Queue Length In Stream (pcu)	1	35.67	35.26	104.00
	2	83.44	83.91	40.98
Delay Per Unit Time In Stream (pcu)	1	30.70	30.46	63.40
	2	46.54	46.77	26.71
Total Delay (pcu·mins)		2317.4	2316.9	2703.3
Reserve Capacity		-26.56%	-26.67%	-16.00%

All the three methods give a maximum cycle time of 120 seconds. As expected, the SIMPLE and NEW methods give less delay, but OSCADY timings makes the junction less oversaturated. From Table 4.8 and Table 4.9 it can also be seen that the NEW method gives fractionally less delay and makes the junction a little more oversaturated than the SIMPLE method. Whereas for the symmetrical junction the differences between the performance with OSCADY timings and with those given by the other two methods were small, in the asymmetrical case they are substantial. The signal settings given by SIMPLE and NEW methods tend to favour the wider road, which becomes less overloaded and has shorter final queue lengths. However, the OSCADY method equalises the degrees of saturation between streams 1 and 2, and gives a longer queue in the wider road.

§4.6.4 Junction 3 — Chapel Hill Junction

This is a real junction in Chapel Hill (Huddersfield, England). It has 4 arms, and has 8 vehicular streams and a pedestrian stream. The delay to the pedestrians is not considered in the optimisation process in this thesis. This junction was first used by **Tully** (1976), as an example junction to illustrate the generation of stage sequences and the calculation of signal timings using **Allsop's** stage-based method. Later **Heydecker** and **Dudgeon** (1987) use this junction to calculating signal settings using the phase-based method. The data for saturation flows and arrival rates (when $X_j<1$, $j=1,2,\cdots,9$), and the junction diagram (Fig 4.7a) and stage diagram (Fig 4.7b) are based on that in **Heydecker** and **Dudgeon** (1987). However, when some $X_j>1$ ($j=1,2,\cdots,9$), the arrival rates are artificial. The Stage information is listed below in Table 4.10.

Table 4.10

Stage	1	2	3	4
Minimum Green (Seconds)	*6.00*	*6.00*	*6.00*	*6.00*
Lost Time After Stage (Seconds)	*5.00*	*1.50*	*5.00*	*5.00*

Stream data:

Table 4.11

	Stream Number								
	1	2	3	4	5	6	7	8	9
Saturation Flow (pcu/h)	*3763*	*3997*	*2622*	*3494*	*2978*	*1835*	*3360*	*2965*	*9000*
Extra Effective Green Time (Seconds)	*7.00*	*0.00*	*0.00*	*6.50*	*0.00*	*0.00*	*1.30*	*0.50*	*0.00*
First Stage In Which Stream Receives Green	*4*	*3*	*4*	*2*	*4*	*4*	*2*	*1*	*3*
Last Stage In Which Stream Receives Green	*4*	*3*	*2*	*2*	*1*	*2*	*3*	*1*	*3*

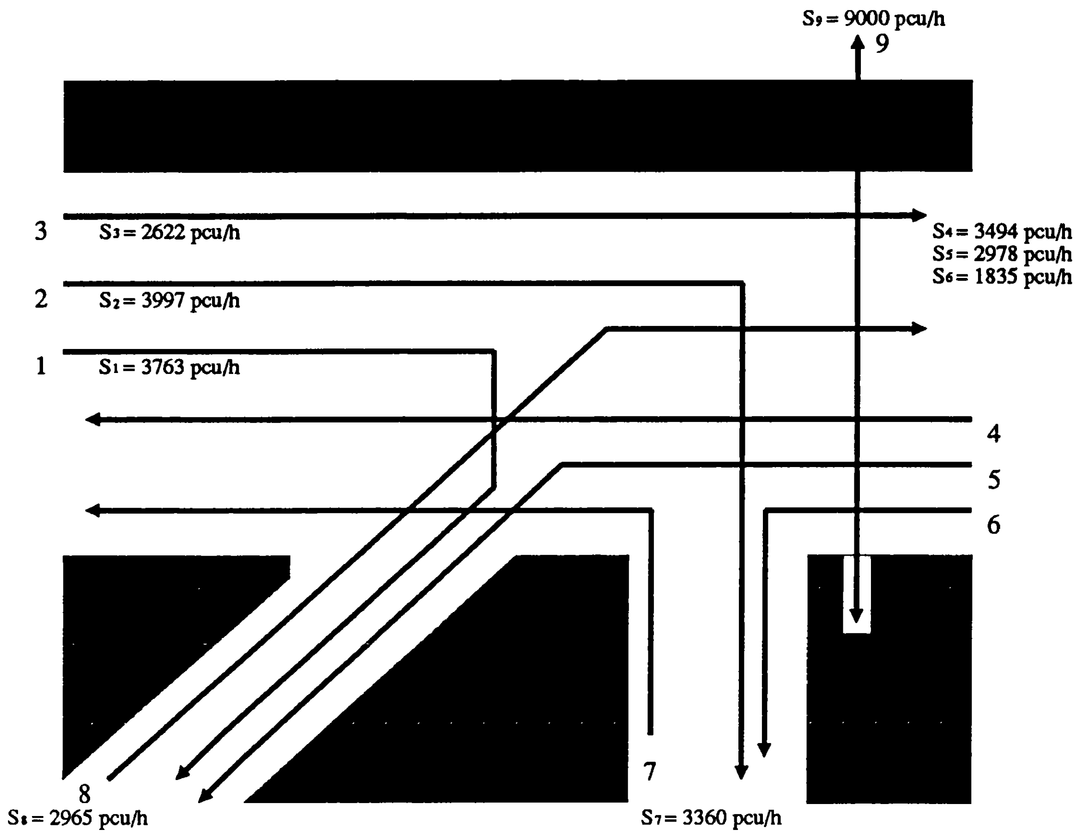


Fig 4.7a Junction 3 - the Chapel Hill junction

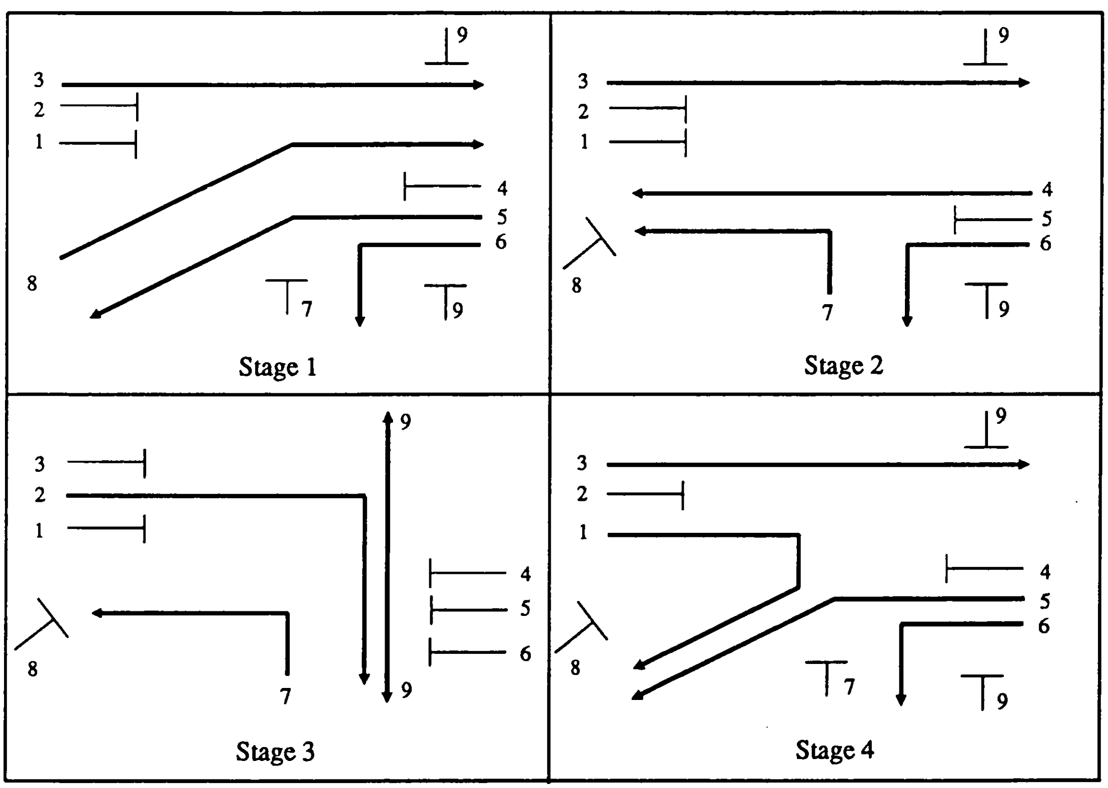


Fig 4.7b The Stage Diagram

Case 1: each $X_j < 1$

The arrival rate in each stream is as follows:

Table 4.12

	Stream				Number				
	1	2	3	4	5	6	7	8	9
Arrival Rate (pcu/h)	*123*	*260*	*250*	*633*	*871*	*722*	*925*	*655*	*1**

*This small arrival rate is introduced to exclude the influence of the pedestrian stream on the optimisation process.

Table 4.13

METHOD		SIGSET	SIMPLE	NEW	OSCADY
Cycle Time (seconds)		53.71	53.11	52.50	60.10
Proportion Of Cycle Effectively Green for Stage	1	0.2786	0.2761	0.2740	0.2962
	2	0.1907	0.1873	0.1831	0.2246
	3	0.1117	0.1130	0.1143	0.1048
	4	0.1117	0.1130	0.1143	0.0998
Proportion Of Cycle Effectively Green for Stream	1	0.2421	0.2448	0.2476	0.2163
	2	0.1117	0.1130	0.1143	0.1048
	3	0.7673	0.7646	0.7619	0.7870
	4	0.3117	0.3097	0.3069	0.3328
	5	0.4835	0.4832	0.4836	0.4792
	6	0.7673	0.7646	0.7619	0.7870
	7	0.3546	0.3530	0.3507	0.3760
	8	0.2879	0.2855	0.2836	0.3045
	9	0.1117	0.1130	0.1143	0.1048
Degree Of Saturation In Stream (%)	1	13.50	13.35	13.20	15.11
	2	58.23	57.57	56.92	62.05
	3	12.43	12.47	12.51	12.11
	4	58.12	58.50	59.03	54.44
	5	60.50	60.53	60.48	61.03
	6	51.28	51.46	51.64	49.99
	7	77.64	77.99	78.50	73.21
	8	76.72	77.39	77.91	72.55
	9	0.10	0.10	0.01	0.11
Total Queue Length In Stream (pcu)	1	0.56	0.55	0.54	0.67
	2	2.12	2.08	2.04	2.46
	3	0.12	0.12	0.12	0.11
	4	3.21	3.21	3.22	3.26
	5	3.00	2.98	2.95	3.36
	6	0.80	0.81	0.82	0.75
	7	5.55	5.57	5.61	5.33
	8	4.66	4.71	4.75	4.53
	9	0.01	0.01	0.01	0.01
Delay Per Unit Time In Stream (pcu)	1	0.56	0.55	0.54	0.67
	2	2.11	2.07	2.03	2.45
	3	0.12	0.12	0.12	0.11
	4	3.21	3.21	3.21	3.26
	5	3.00	2.98	2.94	3.36
	6	0.80	0.81	0.82	0.75
	7	5.53	5.54	5.58	5.32
	8	4.63	4.68	4.71	4.51
	9	0.01	0.01	0.01	0.01
Total Delay (pcu·minutes)		599.0	598.8	598.9	613.0
Reserve Capacity		15.91%	15.40%	14.66%	22.93%

Table 4.13 suggests that for this complicated situation the SIGSET, SIMPLE and NEW method gives nearly the same signal settings and performance. They each give slightly less delay but noticeably less reserve capacity than OSCADY.

Case 2: Some $X_j \geq 1$

The arrival rate in each stream is as follows:

Table 4.14

	Stream				Number				
	1	2	3	4	5	6	7	8	9
Arrival Rate (pcu/h)	*369*	*1500*	*950*	*1 600*	*1700*	*1250*	*1 000*	*700*	1*

*This small arrival rate is introduced to exclude the influence of the pedestrian stream on the optimisation process.

Table 4.15

METHOD		SIMPLE	NEW	OSCADY
Cycle Time (seconds)		*89.50*	*102.66*	*120.00*
Proportion Of Cycle Effectively Green for Stage	1	*0.2588*	*0.2588*	*0.3225*
	2	*0.2827*	*0.3116*	*0.2450*
	3	*0.2071*	*0.2104*	*0.2450*
	4	*0.0670*	*0.0584*	*0.0500*
Proportion Of Cycle Effectively Green for Stream	1	0.1453	0.1266	0.1083
	2	0.2071	0.2104	0.2450
	3	0.7203	0.7263	0.7008
	4	0.3553	0.3749	0.2992
	5	0.3817	0.3660	0.4142
	6	0.7203	0.7263	0.7008
	7	0.5210	0.5493	0.5133
	8	0.2644	0.2637	0.3267
	9	0.2071	0.2104	0.2450
Degree Of Saturation In Stream (%)	1	67.51	77.43	90.52
	2	181.21	178.38	153.18
	3	50.30	49.89	51.70
	4	128.89	122.14	153.07
	5	149.54	155.98	137.83
	6	94.58	93.79	97.20
	7	57.12	54.18	57.98
	8	89.28	89.53	72.27
	9	0.05	0.05	0.05
Total Queue Length In Stream (pcu)	1	4.54	5.97	9.51
	2	345.00	339.98	273.81
	3	1.75	1.89	2.55
	4	191.30	159.32	290.69
	5	291.52	315.97	246.92
	6	11.72	11.28	17.63
	7	4.51	4.51	6.10
	8	10.15	11.16	8.04
	9	0.01	0.01	0.01
Delay Per Unit Time In Stream (pcu)	1	4.53	5.91	8.94
	2	176.94	175.00	143.61
	3	1.75	1.89	2.55
	4	101.61	86.75	152.00
	5	150.72	163.44	130.25
	6	10.66	10.45	15.19
	7	4.51	4.51	6.10
	8	9.81	10.81	8.02
	9	0.01	0.01	0.01
Total Delay (pcu·minutes)		13816.0	13762.0	14000.2
Reserve Capacity		-50.33%	-49.55%	-41.24%

Clearly the three methods give quite different cycle times — the NEW method has a cycle time which is about 13 seconds longer than that by the SIMPLE method and about 17 seconds shorter than that by OSCADY. The fact that the SIMPLE method gives a shorter cycle time than the NEW method in the overloaded case whereas they give very similar cycle times when there was reserve capacity is probably due to the overestimation of uniform delay by the SIMPLE method in oversaturated streams. The NEW method also gives the least total delay of the three methods, though only marginally less than the SIMPLE method, and makes the junction just slightly less oversaturated than the SIMPLE method. Because of capacity maximisation, the signal settings given by OSCADY make the junction least oversaturated, by a margin of 8-9 per cent of the given arrival rates, but as a result the delay is a little greater. The balance between streams is also substantially different, with OSCADY favouring streams 2 and 5 at the expense of stream 4.

4.7 CONCLUSION

These examples suggest that in the simple cases, i.e. for crossroads with simple stage structure, the NEW and SIMPLE methods give similar signal timings, i.e. less delay and just slightly less reserve capacity and noticeably shorter cycle time compared with the SIGSET and OSCADY methods when the junction is undersaturated or oversaturated and symmetrical. When the junction is oversaturated and asymmetrical, however, they give appreciably less delay and more overloading. However, when those methods are applied to the more complicated Chapel Hill Junction, when the junction is undersaturated, the NEW and SIMPLE methods give similar signal timings and similar junction performance to the SIGSET method and give a shorter cycle time and somewhat less delay than OSCADY. When the junction is overloaded, similar results were found for the NEW and SIMPLE methods relative to OSCADY. In both cases, OSCADY gives noticeably more capacity. Hence the new approach is shown to be a practicable alternative to the existing methods, giving results similar to the best of these, and it has the advantage of not relying on an incorrect delay term for oversaturated streams.

In the next chapter, the NEW method will be used to generate optimal signal timings for each of a successive of time periods. These three example junctions will be used again to investigate the problem of settings traffic signals for periods of time-varying demand.

CHAPTER 5. TRAFFIC SIGNAL SETTINGS FOR PERIODS OF TIME-VARYING DEMAND

§5.1 INTRODUCTION

In last chapter the problem of setting traffic signals for an isolated junction during a single time period with steady demand was discussed. As mentioned earlier, the conventional methods of calculating traffic signal settings (e.g. **WEBSTER** 1958, **MILLER** 1963) always assume that the demand flow is time-stationary, and usually one hour is the basic period for calculations. However, practical observations show that traffic demand is time-varying; it can change substantially even within one hour, particularly during peak periods. The way in which flow varies in time has a great influence on the vehicular delays in practical operations and this is reflected in theoretical analyses. Hence it is important to consider the flow change pattern. In most cases the one hour peak period is divided into a sequence of sub-periods (usually 5 — 15 minutes) during each of which the demand flow for each stream is assumed to be constant. The use of time periods shorter than 5 minutes is undesirable because it is difficult to model the flow profile to that detailed level and to evaluate the true demand trends; on the other hand, longer time periods than 15 minutes should be avoided otherwise real flow changes could be missed. Outside peak periods, however, these sub-periods may be longer since the traffic is then usually more stable.

The literature on setting traffic signals for periods of time-varying demand is rather limited. In fact, the only example so far available is the program OSCADY, which can model such a situation, but it gives the signal settings which are in certain cases only local optimal solutions, and optimises with respect to capacity instead of delay during periods of overload. Hence new development is needed to tackle this problem.

In this Chapter, some basic definitions and notation are first introduced, followed by mathematical formulation of the problem. A sequential optimisation method, using the total delay over the whole time period as the performance index, is then established, which adapts the OPTIM subroutine in SIGSET into the process of optimising the effective green times and cycle times in different periods. To prevent longer queue lengths at the end of the last period which will cause unnecessarily bigger delays to the subsequent periods that may follow, the optimiser can re-optimise the objective function by extending the length of the last period so that the effects of the final queues are taken into account. The example calculations show that the new approach can give somewhat better results than OSCADY.

§5.2 STATEMENT OF THE PROBLEM

Fixed-time signal control methods are based on the historical data such as arrival rates, saturation flow rates obtained from traffic survey and traffic counts. A typical pattern of changing flow in one stream over a peak period at a junction is described in Fig 5.1, where T is the length of the peak period, and T_1, T_2, $\cdots$, T_p = (T_k) are the time segments during each of which the flow can be regarded as constant.

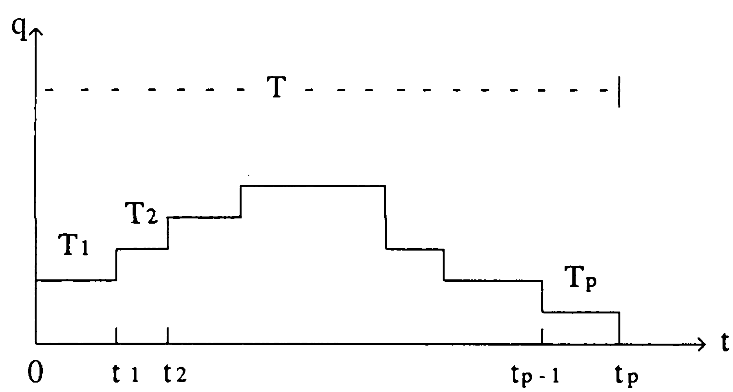


Fig 5.1

A straightforward method of dealing with this kind of problem is to set the traffic signals correspondingly with the variation of traffic flow levels, i.e. to find the optimal signal settings that give the best value of a performance index for each time period. The random queue length of each stream at the end of each time period is treated as the initial random queue length for that stream in the next period. An optimisation technique can then be employed to optimise the traffic signal settings for each single time period, thus a sequence of signal settings $\underline{\lambda_1}$, $\underline{\lambda_2}, \cdots$, $\underline{\lambda_p}$ can be obtained. The program OSCADY is a example of such a method.

However, due to the substantial changes in these traffic characteristics at different times of day, the control strategy should be such that the signal settings can give the best junction performance index taking into account the effects of timings in any one period and conditions in subsequent periods. Since the random queue length for each stream at the end of one segment will be the initial random queue length for that stream in the next segment, the signal settings for each segment are not independent of each other. A control policy that treats them as such will not be optimal when considering different time segments all together and can only be locally optimal in each time interval. There are some ways of making improvements such as vehicle-actuated control methods that will change the traffic signal settings according to the change of flow conditions. However, to investigate the potential use of fixed-time traffic signals, improvements may be made if the signal settings are re-adjusted and subject to time shifts, i.e. if the cycle time and green splits are re-adjusted for each segment but the times at which signal settings are changed for each segment are different from those at which the flow changes, as illustrated in Fig 5.2. Suppose that $(\underline{\lambda_1}', \underline{\lambda_2}', \cdots, \underline{\lambda_p}') = \{\underline{\lambda_k}'\}$ are the new signal settings after re-adjustment that are in force for each period $(T_1', T_2', \cdots, T_p') = \{T_k'\}$ respectively.

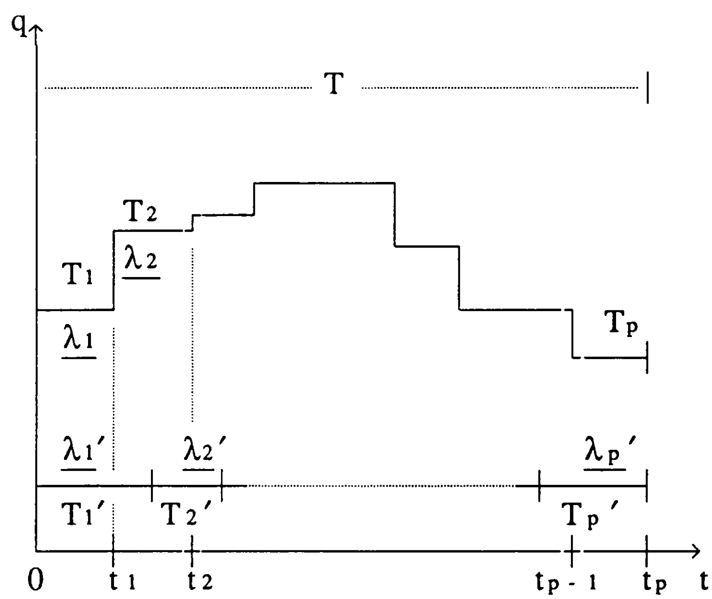


Fig 5.2

Now we have two sequence of time-periods: the one corresponding to the flow changes $\{T_k\} = (T_1, T_2, \cdots, T_p)$ and the one corresponding to the signal settings $\{T_k'\} = (T_1', T_2', \cdots, T_p')$. For the purpose of practical operations let $p' = p$, i.e. the number of changes in signal settings is the same as that in flows. The optimisation problem now becomes how to find the signal settings $(\underline{\lambda}_1', \underline{\lambda}_2', \cdots, \underline{\lambda}_p') = \{\underline{\lambda}_k'\}$ and the sequence of time-periods $\{T_k'\}$ that gives a optimal performance index subject to suitable constraints.

§5.3 NOTATION

Since considerable number of mathematical expressions and operations are involved in this chapter, it is necessary to clarify the notation before further discussion. The following notation will be used in this part of the thesis.

Time Periods

(1) Whole Time: $T\ (t_0,\ t_p)$ (T is divided into p time periods)

(2) The End Of The kth Time Period: t_k $(k = 1, 2, \cdots, p)$

(3) The Length Of The kth Time Period: T_k

The symbol T_k will be used to denote both time period and its duration, so that $T_k = t_k - t_{k-1}$.

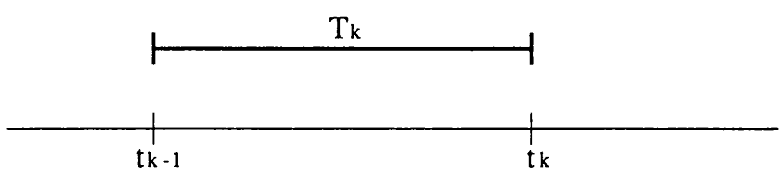


Fig 5.3

(4) Array of p Time Periods

array of p time periods of demand $T_1, T_2, \cdots, T_p$: $\{T_k\}$
array of p time periods of signal settings $T_1', T_2', \cdots, T_p'$: $\{T_k'\}$

$$\{T_k\} = (T_1, T_2, \cdots, T_p)$$
$$\{T_k'\} = (T_1', T_2', \cdots, T_p')$$

Time-shifts

(1) The Shift Of Time T_k: δt_k

δt_k is defined as the difference between the end of T_k and that of T_k', i.e. the amount by which the kth signal settings $\underline{\lambda_k'}$ end ealier or later than $\underline{\lambda_k}$.

$$\delta t_k = t_k' - t_k \quad \begin{cases} \geq 0, & \text{if } t_k' \geq t_k \\ < 0, & \text{if } t_k' < t_k \end{cases}$$

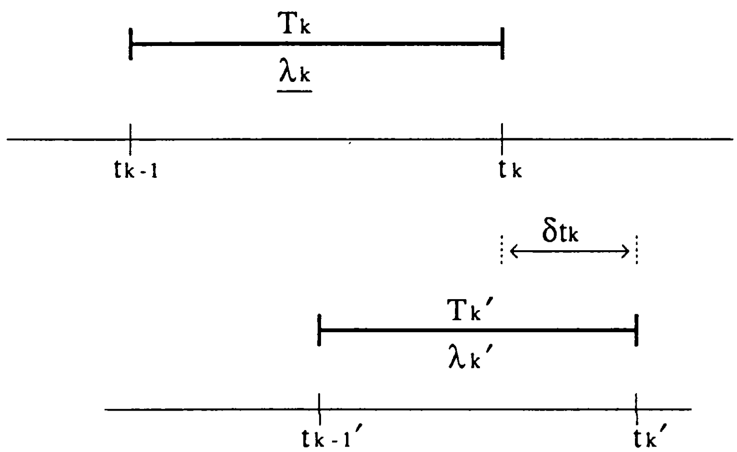


Fig 5.4

(2) Array For Time-shifts In p Time Periods: $\{\delta t_k\}$

$\{\delta t_k\} = (\delta t_0, \ \delta t_1, \ \cdots, \ \delta t_p)$, where $\delta t_0 = \delta t_p = 0$.

Flow In A Stream

flow in stream j: q_j

flow in stream j during T_k: q_{jk}

Saturation Flow In A Stream

saturation flow in stream j: s_j

saturation flow in stream j during T_k: s_{jk}

Traffic Signal Settings

(1) Cycle Time: c

cycle time during T_k : c_k
cycle time during T_k': c_k'

(2) Effective Green Time For A Stage

effective green time for stage i in period T_k: $\lambda_{ik} \cdot c_k$ $i=1, 2, \cdots, m$
effective green time for stage i in period T_k':$\lambda_{ik}' \cdot c_k'$ $i=1, 2, \cdots, m$

(3) Effective Green Time For A Stream

proportion of effective green time for stream j: Λ_j

effective green time for stream j during T_k : $\Lambda_{jk} \cdot c_k$
effective green time for stream j during T_k': $\Lambda_{jk}' \cdot c_k$

(4) Array of Traffic Signal Settings

vector for traffic signal settings for a single period T_k : $\underline{\lambda_k}$
vector for traffic signal settings for a single period T_k': $\underline{\lambda_k'}$

$$\underline{\lambda_k}=(\lambda_{0k}, \lambda_{1k}, \cdots, \lambda_{mk})$$

$$\underline{\lambda_k}'=(\lambda_{0k}', \lambda_{1k}', \cdots, \lambda_{mk}')$$

array for traffic signal settings for p time periods $\{T_k\}$: $\{\underline{\lambda_k}\}$
array for traffic signal settings for p time periods $\{T_k'\}$: $\{\underline{\lambda_k'}\}$

$$\{\underline{\lambda_k}\} = (\underline{\lambda_1}, \underline{\lambda_2}, \cdots, \underline{\lambda_p})$$

$$\{\underline{\lambda_k'}\} = (\underline{\lambda_1'}, \underline{\lambda_2'}, \cdots, \underline{\lambda_p'})$$

Queue Lengths For Stream j: L_j

(1) Uniform Queue Length: L_{uj}

average uniform queue length during the time period T_k: L_{ujk}
average uniform queue length during the time period T_k': L_{ujk}'

(2) Random Queue Length: L_{rj}

random queue length at time t_k: L_{rjk}
random queue length at time t_k': L_{rjk}'

Then L_{rjk} is the initial random queue length for period T_{k+1} and the final random queue length for period T_k, and similarly L_{rjk}' for T_{k+1}' and T_k'.

(3) Total Queue Length for Stream j at t_k: L_{jk}

Total Queue Length for Stream j at t_k': L_{jk}'

$$L_{jk} = L_{rjk} + L_{ujk}$$

Therefore L_{jk} is the initial total queue length for period T_{k+1} and the final total queue length for period T_k, and similarly L_{jk}' for T_{k+1}' and T_k'.

Delays

(1) Delay per unit time: D

delay per unit time in stream j: D_j

If the signal settings are the same throughout T_k it is useful to define

delay per unit time during T_k in stream j: D_{jk}

And if the arrival rates are the same throughout T_k' it is useful to define:

delay per unit time during T_k' in stream j: D_{jk}'

(2) Total delay at the junction: W

total delay at the junction over T_k: W_k

If the signal settings are the same throughout T_k, then

$$W_k = T_k \cdot \sum_{j=1}^{n} D_{jk}$$

total delay at the junction over T_k': W_k'

If the arrival rates are the same throughout T_k', then

$$W_k' = T_k' \cdot \sum_{j=1}^{n} D_{jk}'$$

Where T_k and $T'_{k'}$ overlap, total delay at the junction over $T_k \cap T'_{k'} = W_{kk'}$

(3) Total delay at the junction over the whole time (p time periods):

$$W = \sum_{k=1}^{p} W_k = \sum_{k=1}^{p} W_k'$$

§5.4 MATHEMATICAL FORMULATION

§5.4.1 Basic Assumptions On Signal Settings

In this chapter the following assumptions are made:

1. There are p time periods.

2. The traffic conditions (e.g. flow and saturation flow) in each time period are constant;

3. There are n streams in the junction and m stages in the signal cycle.

4. The constraints on the signal timings (i.e. the stage order, the stage matrix and the minimum green time for each stage, etc) are the same in all the time periods.

Under these assumptions, traffic signal settings for an isolated junction are to be calculated by a stage-based method.

§5.4.2 The Calculation Of The Performance Index

In the following discussions we use the total delay over all the time periods for the junction as a performance index, and the problem can then be expressed as:

To find the signal sequence $(\underline{\lambda_1}', \underline{\lambda_2}', \cdots, \underline{\lambda_p}') = \{\underline{\lambda_k}'\}$ and a series of time periods $(T_1', T_2', \cdots, T_p') = \{T_k'\}$ to minimise

$$W = \sum_{k=1}^{p} W_k' \tag{5.1}$$

where W_k' = total delay in time period T_k' $(k=1,2,\cdots,p)$, subject to the constraints

$$\delta t_1 = 0 \tag{5.2}$$

$$-T_k \le \delta t_k \le T_{k+1} \qquad (k=2, 3, \cdots, p) \tag{5.3}$$

and constraints (2.21 - 2.28).

The constraints (5.3) are introduced to make sure that a certain set of signal settings $\underline{\lambda_k}$ will be implemented in no more than three time periods, i.e. only in extreme cases will $\underline{\lambda_k}$ be implemented in T_{k-1}, T_k and T_{k+1}. Under such constraints the following relationship exists:

$$T_k' = T_k - \delta t_{k-1} + \delta t_k \qquad (k=1,2,\cdots,p) \tag{5.4}$$

By the relationship (5.4), the problem can be restated as follows:

To find the succession of signal settings $(\underline{\lambda_1}', \underline{\lambda_2}', \cdots, \underline{\lambda_p}') = \{\underline{\lambda_k}'\}$ and time-shifts $(\delta t_1, \delta t_2, \cdots, \delta t_p) = \{\delta t_k\}$ to minimise W as defined in (5.1), subject to the constraints (5.2) , (5.3) and (2.21)-(2.28).

The initial settings $(\underline{\lambda}_1, \underline{\lambda}_2, \cdots, \underline{\lambda}_p) = \{\underline{\lambda}_k\}$ are such that $\{T_k'\} = \{T_k\}$, i.e. the initial settings correspond to the flow level in each time segment. The signal settings obtained through the method using the new sheared delay formula introduced in Chapter 4 will be used as such settings. The performance index W corresponding to $\{\underline{\lambda}_k\}$, when $\{T_k'\} = \{T_k\}$, can be easily calculated by:

$$W = \sum_{k=1}^{p} W_k = \sum_{k=1}^{p} \left\{ T_k \sum_{j=1}^{n} D_{jk} \right\} \tag{5.5}$$

where $\quad D_{jk} = D_{ujk} + D_{rjk}$

and $\quad D_{ujk} = D_{ujk}(q_{jk}, \Lambda_{jk}, c_k)$

and according to the sheared delay formula,

$$D_{rjk} = D_{rjk}(q_{jk}, \Lambda_{jk}, L_{rj(k-1)})$$

Suppose, that, through some optimisation techniques, we have found the signal settings $(\underline{\lambda}_1', \underline{\lambda}_2', \cdots, \underline{\lambda}_p') = \{\underline{\lambda}_k'\}$ and a series of time periods $(T_1', T_2', \cdots, T_p') = \{T_k'\}$, then the next problem is how to evaluate the performance index W when $\{\underline{\lambda}_k'\}$ are implemented. Obviously W cannot be evaluated according to (5.5), since there can be more than one set of signal settings in each time period in which the flow is constant. The calculation of the performance index must therefore be conducted in more detail, even within one period. In fact, each time period can be sub-divided into up to three time intervals, in each of which there is only one combination of flow levels and signal settings. If the total number of time periods is p ($p \geq 2$), then the delay in each of these time periods should be considered as follows:

1) Intermediate period T_k

No period of this type exists if $p<3$, and period k is this type for

$k=2$,	if $p=3$.
$k=2, \cdots, p-1$,	if $p>3$

a. $\delta t_{k-1} < 0$, $\delta t_k < 0$

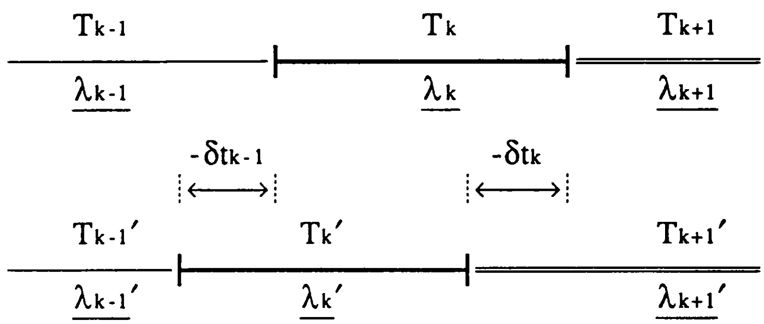


Fig 5.5

There are two sets of signal timings during T_k: $\underline{\lambda_k}'$ and $\underline{\lambda_{k+1}}'$, which are in force for the intervals $T_k \cap T_k'$ and $T_k \cap T_{k+1}'$ respectively. Total delay in the interval $T_k \cap T_k'$ is:

$$W_{kk} = \left\{ \sum_{j=1}^{n} D_j(q_{jk},\ s_{jk},\ \Lambda_{jk}',\ c_k',\ L_{rj(k-1)},\ T_k + \delta t_k) \right\} \cdot (T_k + \delta t_k)$$

The random queue length for stream j at the end of the interval $T_k \cap T_k'$ is:

$$L_{rjk}' = L_{rjk}'(q_{jk},\ s_{jk},\ \Lambda_{jk}',\ L_{rj(k-1)},\ T_k + \delta t_k)$$

Then the total delay in the interval $T_k \cap T_{k+1}'$ is:

$$W_{k(k+1)} = \left\{ \sum_{j=1}^{n} D_j(q_{jk},\ s_{jk},\ \Lambda_{j(k+1)}',\ c_{k+1}',\ L_{rjk}',\ -\delta t_k) \right\} \cdot (-\delta t_k)$$

Hence the total delay in period T_k is:

$$W_k = W_{kk} + W_{k(k+1)}$$

The random queue length for stream j at the end of period T_k is:

$$L_{rjk} = L_{rjk}(q_{jk},\ s_{jk},\ \Lambda_{j(k+1)}',\ L_{rjk}',\ -\delta t_k)$$

b. $\delta t_{k-1} < 0,\ \delta t_k \geq 0$

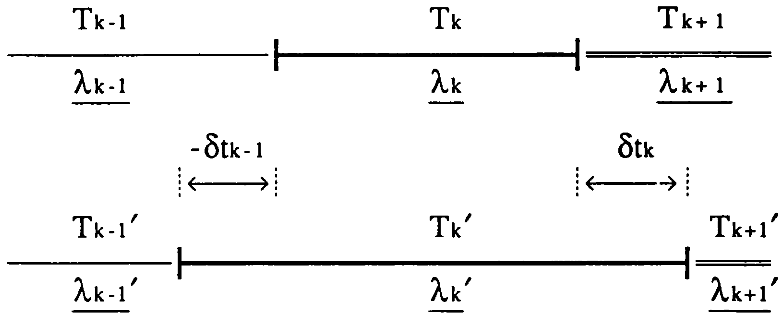


Fig 5.6

In this case there is only one set of signal settings $\underline{\lambda}_k'$ in T_k, hence the total delay in T_k is:

$$W_k = \left\{ \sum_{j=1}^{n} D_j(q_{jk},\ s_{jk},\ \Lambda_{jk}',\ c_k',\ L_{rj(k-1)},\ T_k) \right\} \cdot T_k$$

The random queue length for stream j at the end of T_k is:

$$L_{rjk} = L_{rjk}(q_{jk},\ s_{jk},\ \Lambda_{jk}',\ L_{rj(k-1)},\ T_k)$$

c. $\delta t_{k-1} \geq 0,\ \delta t_k < 0$

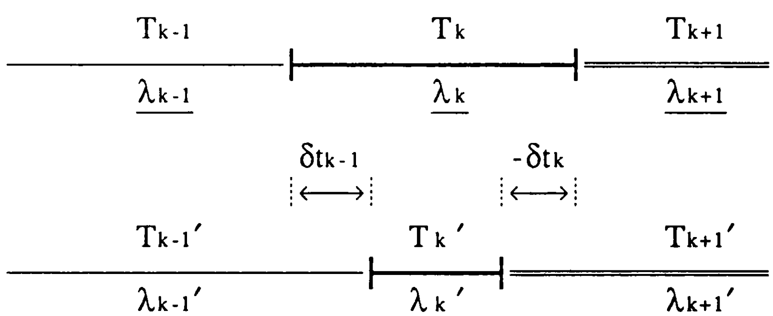


Fig 5.7

There are three different sets of signal settings: $\underline{\lambda}_{k-1}'$, $\underline{\lambda}_k'$ and $\underline{\lambda}_{k+1}'$, which are in force in three intervals: $T_k \cap T_{k-1}'$, T_k' and $T_k \cap T_{k+1}'$ respectively. The total delay in the first interval $T_k \cap T_{k-1}'$ is:

$$W_{k(k-1)} = \left\{ \sum_{j=1}^{n} D_j(q_{jk},\ s_{jk},\ \Lambda_{j(k-1)}',\ c_{k-1}',\ L_{rj(k-1)},\ \delta t_{k-1}) \right\} \cdot \delta t_{k-1}$$

The random queue length for stream j at the end of this interval is:

$$L_{rj(k-1)}' = L_{rj(k-1)}'(q_{jk},\ s_{jk},\ \Lambda_{j(k-1)}',\ L_{rj(k-1)},\ \delta t_{k-1})$$

Then the total delay in the second interval T_k' is:

$$W_{kk} = \left\{\sum_{j=1}^{n} D_j(q_{jk},s_{jk},\Lambda_{jk}',c_k',L_{rj(k-1)}',T_k-\delta t_{k-1}+\delta t_k)\right\}\cdot(T_k-\delta t_{k-1}+\delta t_k)$$

The random queue length for stream j at the end of this interval is:

$$L_{rjk}' = L_{rjk}'(q_{jk},\ s_{jk},\ \Lambda_{jk}',\ L_{rj(k-1)}',\ T_k-\delta t_{k-1}+\delta t_k)$$

Hence the total delay in the interval $T_k \cap T_{k+1}'$ is:

$$W_{k(K+1)} = \left\{\sum_{j=1}^{n} D_j(q_{jk},s_{jk},\Lambda_{j(k+1)}',c_{k+1}',L_{rjk}',-\delta t_k)\right\}\cdot(-\delta t_k)$$

Therefore the total delay in period T_k can be given by:

$$W_k = W_{k(k-1)} + W_{kk} + W_{k(K+1)}$$

The random queue length for stream j at the end of period T_k is:

$$L_{rjk} = L_{rjk}(q_{jk},\ s_{jk},\ \Lambda_{j(k+1)}',\ L_{rjk}',\ -\delta t_k)$$

d. $\delta t_{k-1} \geq 0$, $\delta t_k \geq 0$

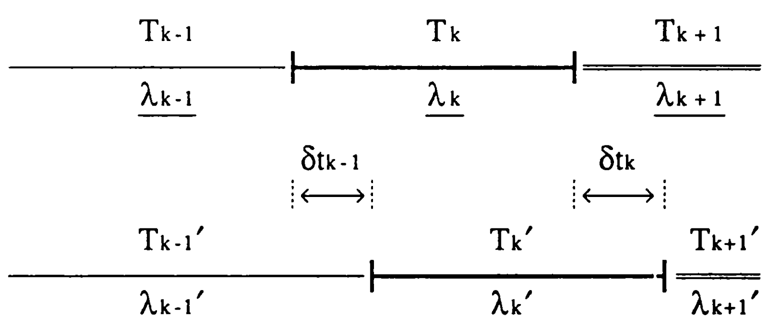


Fig 5.8

As in case a), there are two sets of signal timings $\underline{\lambda_{k-1}'}$ and $\underline{\lambda_k'}$ in the intervals $T_k \cap T_{k-1}'$ and $T_k \cap T_k'$ respectively. The total delay in the first interval $T_k \cap T_{k-1}'$ is:

$$W_{k(k-1)} = \left\{ \sum_{j=1}^{n} D_j(q_{jk},\ s_{jk},\ \Lambda_{j(k-1)}',\ c_{k-1}',\ L_{rj(k-1)},\ \delta t_{k-1}) \right\} \cdot \delta t_{k-1}$$

The random queue length for stream j at the end of this interval is:

$$L_{rj(k-1)}' = L_{rj(k-1)}'(q_{jk},\ s_{jk},\ \Lambda_{j(k-1)}',\ L_{rj(k-1)},\ \delta t_{k-1})$$

Then the total delay in the second interval $T_k \cap T_k'$ is:

$$W_{kk} = \left\{ \sum_{j=1}^{n} D_j(q_{jk}, s_{jk}, \Lambda_{jk}', c_k', L_{rj(k-1)}', T_k - \delta t_{k-1}) \right\} \cdot (T_k - \delta t_{k-1})$$

Therefore the total delay in period T_k can be given by:

$$W_k = W_{k(k-1)} + W_{kk}$$

The random queue length for stream j at the end of period T_k is:

$$L_{rjk} = L_{rjk}(q_{jk},\ s_{jk},\ \Lambda_{jk}',\ L_{rj(k-1)}',\ T_k - \delta t_{k-1})$$

2) First period T_1

By definition $\delta t_0 = 0$.

a. If $\delta t_1 < 0$.

The total delay in T_1 can be calculated by putting k=1 and $\delta t_{k-1}=0$ in case (c) for the intermediate period.

b. If $\delta t_1 \geq 0$.

The total delay in T_1 can be calculated by putting k=1 and $\delta t_{k-1}=0$ in case (d) for the intermediate period.

3) Last period T_p

By definition $\delta t_p = 0$.

a. If $\delta t_{p-1} < 0$

The total delay in T_p can be calculated by putting k=p and $\delta t_k=0$ in case (b) for the intermediate period.

b. If $\delta t_{p-1} \geq 0$

The total delay in T_p can be calculated by putting k=p and $\delta t_k=0$ in case (d) for the intermediate period.

§5.4.3 Evaluation Of The Effects Of $\{\underline{\lambda_k}'\}$ and $\{\delta t_k\}$ On The Peformance Index

After the two performance indices W and W′ corresponding to two successions of signal settings $\{\underline{\lambda_k}\}$ and $\{\underline{\lambda_k}'\}$ have been calculated, it is important to compare W and W′ so that the effects of re-adjusting the signal timings and making time-shifts can be evaluated. The criterion that will be used for this purpose is the percentage reduction in total delay:

$$\delta W/W = (W-W')/W \tag{5.6}$$

The criterion $\delta W/W$ will be used in example calculations.

§5.5 OPTIMISATION OF $\{\underline{\lambda}_k'\}$ FOR FIXED $\{\delta t_k\}$

§5.5.1 Introduction

The discussions in this Chapter have not yet considered the problem of how to find $\{\underline{\lambda}_k'\}$ and $\{\delta t_k'\}$, but only the evaluation of the Performance Index once $\{\underline{\lambda}_k'\}$ and $\{\delta t_k'\}$ are found. In fact, it is difficult to solve the problem of simultaneous optimisation of $\{\underline{\lambda}_k\}$ and $\{\delta t_k\}$, since the number of variables involved is very large and no existing technique is available except exhaustive grid search. However, the problem can be solved by the sequential optimisation of $\{\underline{\lambda}_k'\}$ and $\{\delta t_k\}$ iteratively. This section introduces the optimisation of $\{\underline{\lambda}_k'\}$ without changing $\{\delta t_k\}$, and the next section §5.6 discusses the problem of optimising $\{\delta t_k\}$ under fixed $\{\underline{\lambda}_k'\}$. An iterative algorithm of sequential optimising $\{\underline{\lambda}_k'\}$ and $\{\delta t_k\}$ is given in §5.7.

When $\{\delta t_k\}$ are fixed, to optimise $\{\underline{\lambda}_k'\}$, we must first consider the factors affecting the choice of $\{\underline{\lambda}_k'\}$. First consider the problem of optimising a particular $\underline{\lambda}_k'$, $k = 1, 2, \cdots, p$. Since the random queue lengths at the end of T_k' will be the initial queue lengths at the beginning of T_{k+1}', any change in $\underline{\lambda}_k'$ will directly influence the delay in the subsequent periods after T_k'. On the other hand, such a change will also indirectly influence the delay in the periods before T_k', since the change in $\underline{\lambda}_k'$ will eventually cause changes in $\underline{\lambda}_h'$ and hence W_h', $h = 1, 2, \cdots, k-1$. Therefore the signal settings in one period will be determined not only by the traffic conditions in this period, but also by the traffic conditions and signal settings in the periods before and after this period, i.e. $\underline{\lambda}_k'$ must be treated as global variables, and such influences must be considered in order to get the global optimal signal settings.

§5.5.2 The Calculation Of The Derivatives With Respect To $\underline{\lambda}_k'$

To optimise $\underline{\lambda}_k'$ $(k=1, 2, \cdots, p)$, it is important to evaluate the derivatives of W with respect to $\underline{\lambda}_k'$, i.e. $\partial W/\partial \lambda_{ik}'$ $(i=1,2,\cdots,m,$ $k=1,2,\cdots,p)$ should be known. At each step in the iterative process of optimisation to be introduced in §5.5.3 below, we optimise $\underline{\lambda}_k'$ under existing $\underline{\lambda}_h'$, $h = 1, 2,\cdots,k-1, k+1,\cdots, p$. In doing so we only consider the direct effect of changing $\underline{\lambda}_k'$ on the delay in periods after T_k', and the indirect effect of changing $\underline{\lambda}_k'$ on the delay in periods before T_k' is taken into account later in the iterative optimisation process.

$$\text{Now} \quad W = \sum_{j=1}^{n} \left\{ \sum_{h=1}^{p} W_{jh}' \right\} \tag{5.7}$$

Where W_{jh}' is the delay in stream j during T_h', and the change in $\underline{\lambda}_k'$ will only have direct effect on W_h', h=k,k+1,···,p. W_k' itself is influenced through changes in the Λ_{jk}' (j=1,2,···,n) (in accordance with the relationship (2.21)); and the W_h' (h=k+1, k+2, ···, p) are influenced through changes in the initial queue lengths $L_{rj(h-1)}'$ (j = 1, 2, ···, n; h = k+1, k+2, ···, p).

Hence

$$\frac{\partial W}{\partial \lambda_{ik}'} = \sum_{j=1}^{n} \left\{ \sum_{h=1}^{p} \frac{\partial W_{jh}'}{\partial \lambda_{ik}'} \right\} = \sum_{j=1}^{n} \left\{ \sum_{h=k}^{p} \frac{\partial W_{jh}'}{\partial \Lambda_{jk}'} \cdot \frac{\partial \Lambda_{jk}'}{\partial \lambda_{ik}'} \right\} = \sum_{j=1}^{n} a_{ij} \left\{ \sum_{h=k}^{p} \frac{\partial W_{jh}'}{\partial \Lambda_{jk}'} \right\}$$

$$= \sum_{j=1}^{n} a_{ij} \left\{ \frac{\partial W_{jk}'}{\partial \Lambda_{jk}'} + \frac{\partial L_{rjk}'}{\partial \Lambda_{jk}'} \left(\frac{\partial W_{j(k+1)}'}{\partial L_{rjk}'} + \sum_{h=k+2}^{p} \left[\frac{\partial W_{jh}'}{\partial L_{rj(h-1)}} \prod_{v=k+2}^{h} \frac{\partial L_{rj(v-1)}'}{\partial L_{rj(v-2)}'} \right] \right) \right\}$$

$$i = 1, 2, \cdots, m \tag{5.8}$$

$$\frac{\partial W}{\partial \lambda_{0k}'} = \sum_{j=1}^{n} \left\{ \sum_{h=1}^{p} \frac{\partial W_{jh}'}{\partial \lambda_{0k}'} \right\} = \sum_{j=1}^{n} \left\{ \sum_{h=k}^{p} \frac{\partial W_{jh}'}{\partial \Lambda_{jk}'} \cdot \frac{\partial \Lambda_{jk}'}{\partial \lambda_{0k}'} + \frac{\partial W_{jk}'}{\partial \lambda_{0k}'} \right\}$$

$$= \sum_{j=1}^{n} \left\{ a_{0j} \left(\sum_{h=k}^{p} \frac{\partial W_{jh}'}{\partial \Lambda_{jk}'} \right) + \frac{\partial W_{jk}'}{\partial \lambda_{0k}'} \right\} \tag{5.9}$$

The derivatives $\partial W_{jk}'/\partial \Lambda_{jk}'$, $\partial W_{jk}'/\partial \lambda_{0k}'$ and $\partial W_{jk}'/\partial L_{rj(k-1)}'$ are multiples of the corresponding derivatives of D_{jk}, which are given in Appendix 2. The derivatives $\partial L_{rjk}'/\partial \Lambda_{jk}'$ and $\partial L_{rjk}'/\partial L_{rj(k-1)}'$ are given in Appendix 3. And when $\{\delta t_k\} \neq \{0\}$, by analogy with the discussion in §5.4.2, there are further steps in the calculation, i.e. the period T_k' may be sub-divided into up to three time-intervals according to the number of different set of demand flow levels. For example, when k=1, to calculate $\partial W_{j1}'/\partial \Lambda_{j1}'$:

a. If $\delta t_1 \le 0$, since there is no change in the flow q_{j1} during T_1', $\partial W_{j1}'/\partial \Lambda_{j1}'$ can be calculated directly by

$$\partial W_{j1}'/\partial \Lambda_{j1}' \quad = \frac{\partial}{\partial \Lambda_{j1}'}\left(W_{j1}'(q_{j1},\ s_{j1},\ \Lambda_{j1}',\ c_1',\ L_{rj0},\ T_1')\right)$$

b. If $\delta t_1 > 0$, since $T_1' = T_1 + \delta t_1$, the flow q_{j1} is changed to q_{j2} at the end of T_1, hence $\partial W_{j1}'/\partial \Lambda_{j1}'$ consists of two components corresponding to periods T_1 and $T_2 \cap T_1'$, and given by

$$\frac{\partial}{\partial \Lambda_{j1}'}\left(W_{j11}(q_{j1},\ s_{j1},\ \Lambda_{j1}',\ c_1',\ L_{rj0},\ T_1)\right)$$

and

$$\frac{\partial}{\partial \Lambda_{j1}'}\left(W_{j21}(q_{j2},\ s_{j2},\ \Lambda_{j1}',\ c_1',\ L_{rj1},\ \delta t_1)\right)$$

$$+ \frac{\partial W_{j21}}{\partial L_{rj1}} \cdot \frac{\partial}{\partial \Lambda_{j1}'}\left(L_{rj1}(q_{j1}, s_{j1}, \Lambda_{j1}', L_{rj0}, T_1)\right)$$

where W_{jkk}' is the delay in stream j during $T_k \cap T_k'$.

Hence $\partial W/\partial \lambda_{ik}'$ $(i=1,2,\cdots,m,\ k=1,2,\cdots,p)$ can be evaluated according to the above formulation using the new sheared delay formula (4.20).

§5.5.3 The Sequential Optimisation Of $\{\underline{\lambda}_k'\}$

To optimise $\underline{\lambda}_k'$, $k = 1, 2, \cdots, p$, the SIGSET optimisation subroutine OPTIM that was used to calculate the optimal signal settings for a single period can still be used here, since the signal timing requirements (e.g. the definition of the stage matrix) and the constraints (2.21)-(2.28) are unchanged and the variables are the same. The only difference is in the objective function, which is reflected in the difference in the derivatives.

Hence $\{\underline{\lambda}_k'\}$ can be optimised for given $\{\delta t_k\}$ by the following sequential algorithm:

Step 0: **ALGORITHM 0** *(Initial Signal Settings)*

Find the initial solution $\{\underline{\lambda}_k'\}=\{\underline{\lambda}_k\}$, where $\{\underline{\lambda}_k\}$ is the sequence of traffic signal settings that give the optimal performance index for each single time period T_k, $k=1, 2, \cdots, p$.

ALGORITHM 1 *(Sequential Optimisation of $\{\underline{\lambda}_k'\}$:*

Choose a small ε_1 (e.g. $\varepsilon_1=0.0001$) for use in the criterion for stopping the algorithm.

Step 1: Let the current Performance Index be W_c. For given $\{\delta t_k'\}$, optimise $(\underline{\lambda}_1', \underline{\lambda}_2', \cdots, \underline{\lambda}_p')=\{\underline{\lambda}_k'\}$.

Step 1.1 Optimise $\underline{\lambda}_1'$.
Step 1.2 Optimise $\underline{\lambda}_2'$.
......
Step 1.p Optimise $\underline{\lambda}_p'$.

Steps 1.1 - 1.p are realised by implementing OPTIM, which uses the derivatives given by (5.8) and (5.9). Then a new value of the Performance Index is obtained, which is denoted by W_{new}.

Step 2. If $(W_c-W_{new}) > \varepsilon_1 \cdot W_c$, go to step 1.1. Otherwise stop.

Hence $\{\underline{\lambda}_k'\}$ is optimised with respect to the total delay W in the whole period T rather than each λ_k' being optimised individually with respect to the total delay W_k' in period T_k'. i.e. the effects of each setting $\underline{\lambda}_k'$ on W are considered.

§5.5.4 An Example Calculation

To illustrate the practicability of the above algorithm, we consider the symmetric crossroads – junction 1, which was introduced in §4.6.2. This example junction will be used throughout this chapter to illustrate the various algorithms. The additional data to Table 4.2 are given in Table 5.1.

Table 5.1 Additional Data For Example Junction 1

	Period T_1	Period T_2
Length Of Time Period	T_1=10 minutes	T_2=10 minutes
Flow Ratio In Stream 1	$y_{11} = 0.6$	$y_{12} = 0.45$
Flow Ratio In Stream 2	$y_{21} = 0.4$	$y_{22} = 0.3$
Initial Queue Length In Stream 1	$L_{r10} = 0.0$	
Initial Queue Length In Stream 2	$L_{r20} = 0.0$	

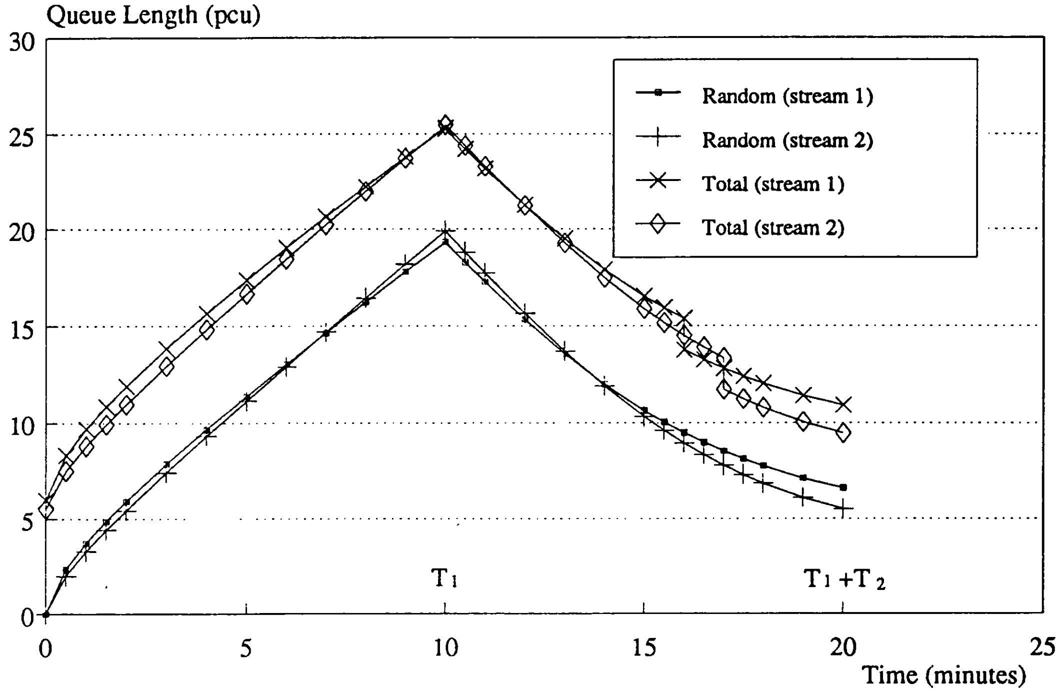
The initial signal settings obtained from Algorithm 0 are summarised in Table 5.2, where the total delay in period 1 and 2, W_1 and W_2, are estimated by the new delay formula (4.20), since this is the formula used in the optimisation process. However, the total delay for the whole period (T_1 and T_2 together) is estimated by both the new delay expression and the OSCADY formula, the results being denoted by W_{new} and W_{osc} respectively. The latter value is calculated to enable subsequent direct comparison with timings calculated by OSCADY. The evolution of the queues is plotted in Fig 5.9.

Table 5.2 The Initial Signal Settings $\{\underline{\lambda_k}\}$ for Example Junction 1

	Period T_1	Period T_2
Cycle Time (Seconds)	$c_1 = 87.49$	$c_2 = 85.38$
Allocations of Green Times $\underline{\lambda_1}, \underline{\lambda_2}$	$\lambda_{11} = 0.5583$ $\lambda_{21} = 0.3502$	$\lambda_{12} = 0.5291$ $\lambda_{22} = 0.3772$
Final Random Queue Length	$L_{r11} = 19.33$ $L_{r21} = 19.91$	$L_{r12} = 6.61$ $L_{r22} = 5.51$
Final Uniform Queue Length	$L_{u11} = 5.99$ $L_{u11} = 5.53$	$L_{u12} = 4.30$ $L_{u22} = 3.94$
Total Delay (pcu·minutes)	W_1= **338.30**	W_2 = **321.85**
	W_{new} = **660.15** W_{osc} = **654.00**	

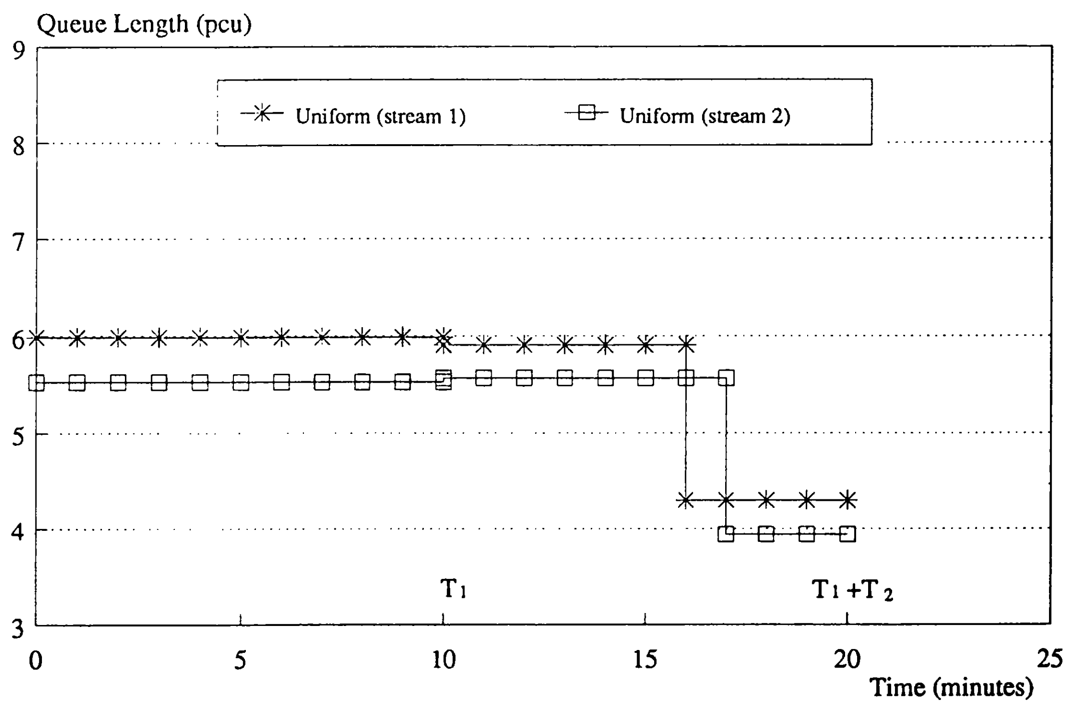
Starting from this initial solution, the signal settings are re-optimised using the Algorithm 1. The results are summarised in Table 5.3. The evolution of the queues in this case can be seen in Fig 5.10.

Fig 5.9 Queue Lengths
With Initial Signal Settings



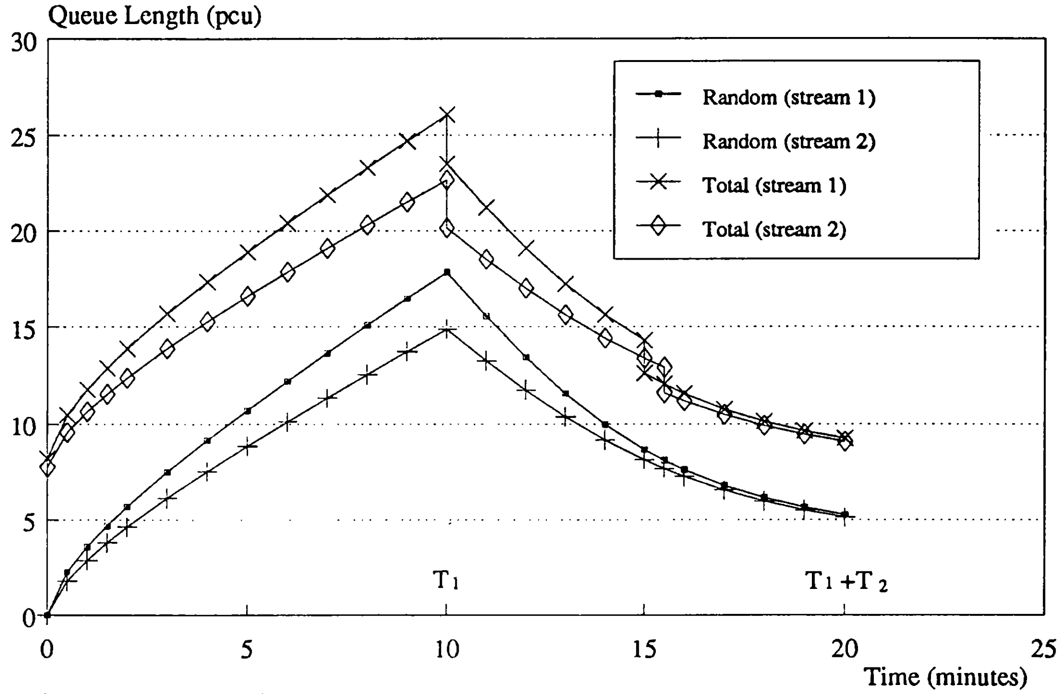
(a) Total and random queue lengths

Fig 5.9 Queue Lengths
With Initial Signal Settings



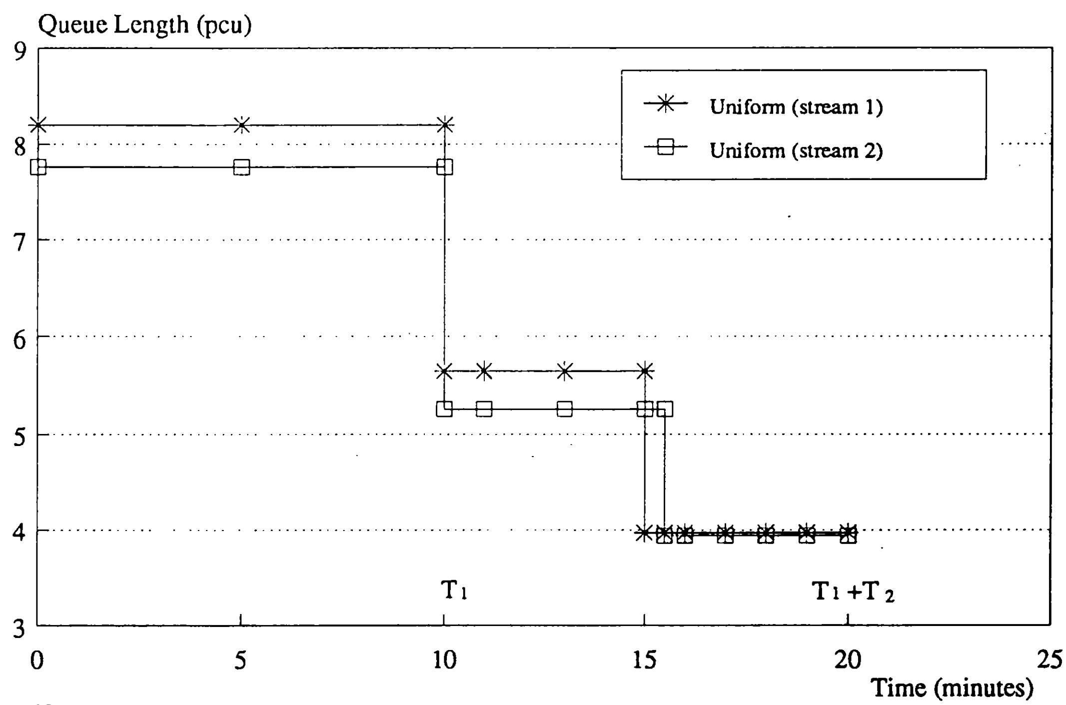
(b) Unifom queue lengths

Fig 5.10 Queue Lengths
With Re-optimised Signal Settings



(a) Total and random queue lengths

Fig 5.10 Queue Lengths
With Re-optimised Signal Settings



(b) Uniform queue lengths

From Table 5.2 and Table 5.3 it can be seen that the initial settings are only a local solution to the problem, since they are only optimal in each single period, and even though the delay in each of these single periods is minimal, the queues at the end of each period are not taken into account, which may cause bigger delay to the subsequent periods than the case where these queues are considered in the optimisation process. In this example the initial settings result in a delay of W_1 = 338.30 pcu·minutes in T_1 and W_2 = 321.85 pcu·minutes in T_2. By implementing Algorithm 1, although the delay in T_1 is increased to W_1 = 350.96 pcu·minutes, the delay in T_2 is substantially reduced to W_2 = 273.11 pcu·minutes, hence the total delay in the whole period, $W = W_1 + W_2$ is less than before, i.e. the overall performance index is improved. This can be explained by comparing Fig 5.9 and Fig 5.10, from which it can be seen that, at the end of T_1, the initial settings give a total queue length of L_{11} = 25.32 pcu for stream 1 and L_{21} = 25.44 pcu for stream 2. However, the re-optimised signal settings give a total queue length of L_{11}=26.05 pcu and L_{21}=22.65 pcu, and although L_{11} becomes bigger, $L_{11}+L_{21}$ becomes smaller. The queues at the end of T_1 are thus shorter in total, so that the delays in T_2 are smaller than before. As a result, an appreciable decrease in total delay of about 5.5% are obtained by the re-optimisation. Moreover, the queues remaining at the end of T_2 are shorter in total after optimisation.

Table 5.3 The Signal Settings After Re-optimising $\{\underline{\lambda_k'}\}$ for Example Junction 1

	Period T_1	Period T_2
Cycle Time (Seconds)	$c_1 = 120.00$	$c_2 = 81.78$
Allocations of Green Times $\underline{\lambda_1}, \underline{\lambda_2}$	$\lambda_{11}' = 0.5642$ $\lambda_{21}' = 0.3691$	$\lambda_{12}' = 0.5380$ $\lambda_{22}' = 0.3641$
Final Random Queue Length	$L_{r11} = 17.85$ $L_{r21} = 14.89$	$L_{r12} = 5.28$ $L_{r22} = 5.12$
Final Uniform Queue Length	$L_{u11} = 8.20$ $L_{u21} = 7.76$	$L_{u12} = 3.97$ $L_{u22} = 3.94$
Total Delay (pcu·minutes)	$W_1 = 350.96$	$W_2 = 273.11$
	$W_{new} = 624.07$ $W_{osc} = 618.47$	
Percentage Reduction* In Total Delay $100 \cdot \delta W / W$	5.47 % by the new formula 5.43 % by the OSCADY formula	

*: compared with initial settings (Table 5.2)

§5.6 OPTIMISATION OF $\{\delta t_k\}$ FOR FIXED $\{\underline{\lambda}_k'\}$

§5.6.1 Introduction

In the above algorithm, although the total delay for the whole periods can be decreased by re-optimising $\{\underline{\lambda}_k'\}$, it may be decreased further if the signal settings are subject to time-shift, i.e. if the change of signal settings from one period to the next is not necessarily simultaneous with the change of the flows but can be earlier or later. To investigate this possibility, calculations are made using the same example as in §5.5, where the initial signal settings $\{\underline{\lambda}_k\}$ and the re-optimised settings $\{\underline{\lambda}_k'\}$ are both tested, i.e. by continuing $\underline{\lambda}_1$ (or $\underline{\lambda}_1'$) into the second period (in which case $\delta t_1>0$), or by starting $\underline{\lambda}_2$ (or $\underline{\lambda}_2'$) before the end of the first period (in which case $\delta t_1<0$).

§5.6.2 The Example Calculations

The effects on the total delay for the whole periods of making the time-shift to the initial settings $\{\underline{\lambda}_k\}$ and the re-optimised settings $\{\underline{\lambda}_k'\}$ can be seen in Fig 5.11a and Fig 5.12a, where the delays are estimated by the new delay expression, and in Fig 5.11b and Fig 5.12b, where the delays are estimated by the OSCADY formula. From these curves it can be seen that, for the initial signal settings $\{\underline{\lambda}_k\}$ the time-shift for minimum total delay is $\delta t_1=36$ seconds, i.e. W is minimised if the signal settings are changed 36 seconds later than the change of flows; and for the re-optimised signal settings $\{\underline{\lambda}_k'\}$ the time-shift which yields minimum total delay is $\delta t_1=72$ seconds.

Figures similar to Fig 5.11 were plotted for the cases listed in Table 5.4 and Table 5.5 (for all these cases, T_1 and T_2 were also reversed), and the following characteristics were found:

1. The curve for the total delay W as a function of the time shift δt_k ($k=1, 2, \cdots, p$), $W = W(\delta t_k)$, is non-convex, and non-differentiable at $\delta t_k = 0$ ($k = 1, 2, \cdots, p-1$) under fixed $\{\underline{\lambda}_k'\}$.

Table 5.4 Cases Studied For Example Junction 1

Period 1 (T_1=10 minutes)		Period 2 (T_2=10 minutes)	
y_{11}	y_{21}	y_{12}	y_{22}
0.6	0.4	0.45	0.3
0.6	0.4	0.3	0.2
0.75	0.5	0.5	0.33
0.75	0.5	0.375	0.25
0.75	0.5	0.3	0.2

Table 5.5 Cases Studied For Example Junction 2

Period 1 (T_1=10 minutes)		Period 2 (T_2=10 minutes)	
y_{11}	y_{21}	y_{12}	y_{22}
0.6	0.4	0.3	0.2
0.6	0.4	0.2	0.3
0.4	0.6	0.2	0.3
0.4	0.6	0.3	0.2

2. W = $W(\delta t_k)$ has only one turning value, a minimum, in a neighbourhood of zero which is very often as large as the largest range $(-T_k, T_{k+1})$ of values of δt_k that can need to be considered within the constraints (5.3). In least favorable cases like this example, there are other turning values towards the extremities of the range. In such cases, as δt_k increases from $-T_k$, $W(\delta t_k)$ may pass through another turning value before decreasing into the neighbourhood of zero that contains a minimum. Again as δt_k increases towards T_{k+1}, $W(\delta t_k)$ may pass through another turning value. In the example shown here, there is one local maximum point for $W(\delta t_k)$ near $\delta t_1 = -T_1 = -10$ minutes.

3. Where there are other local maximum or local minimum points for $W(\delta t_k)$ in the cases studied the δt_k such that $W = W(\delta t_k)$ attains its lowest value is always the one minimum point in the neighbourhood of zero.

Delay As a Function of Time-shift With Initial Signal Settings

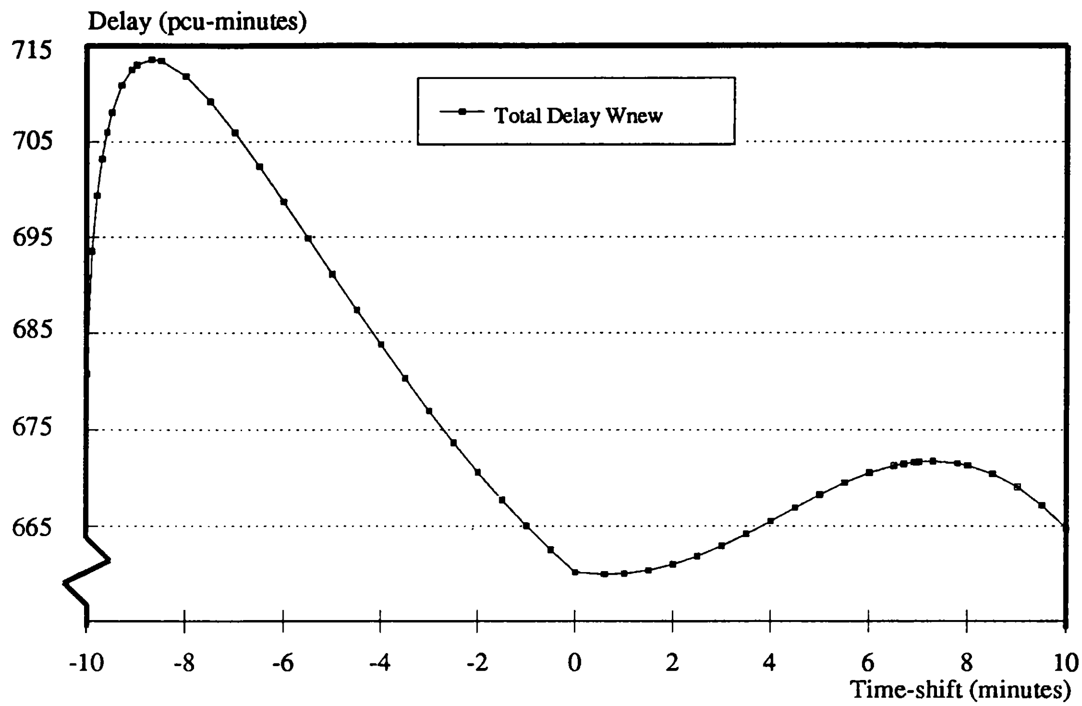


Fig 5.11a Delay Estimated By The New Delay Formula

Delay As a Function of Time-shift With Initial Signal Settings

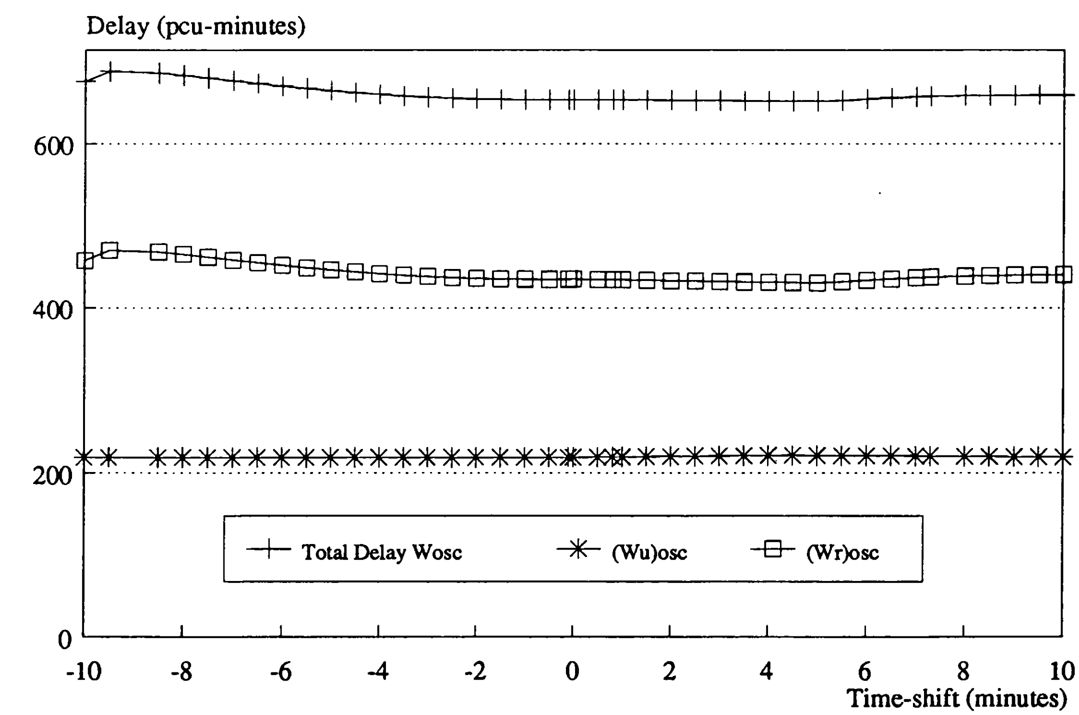


Fig 5.11b Delay Estimated By The OSCADY Delay Formula

Delay As a Function of Time-shift With Re-optimised Signal Settings

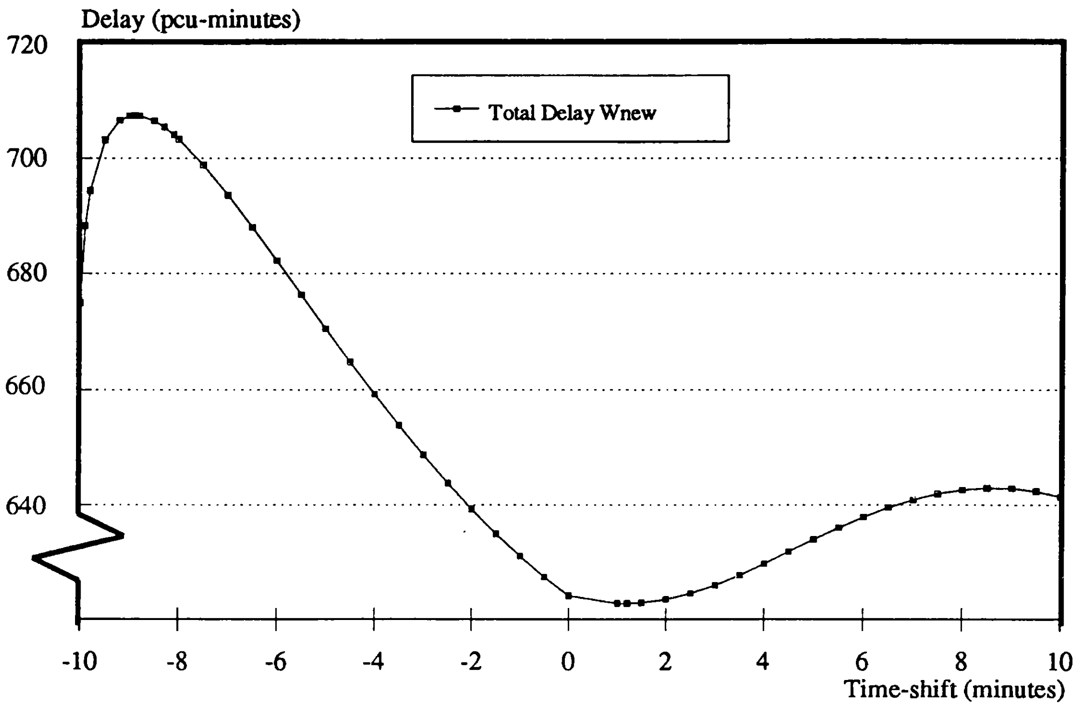


Fig 5.12a Delay Estimated By The New Delay Formula

Delay As a Function of Time-shift With Re-optimised Signal Settings

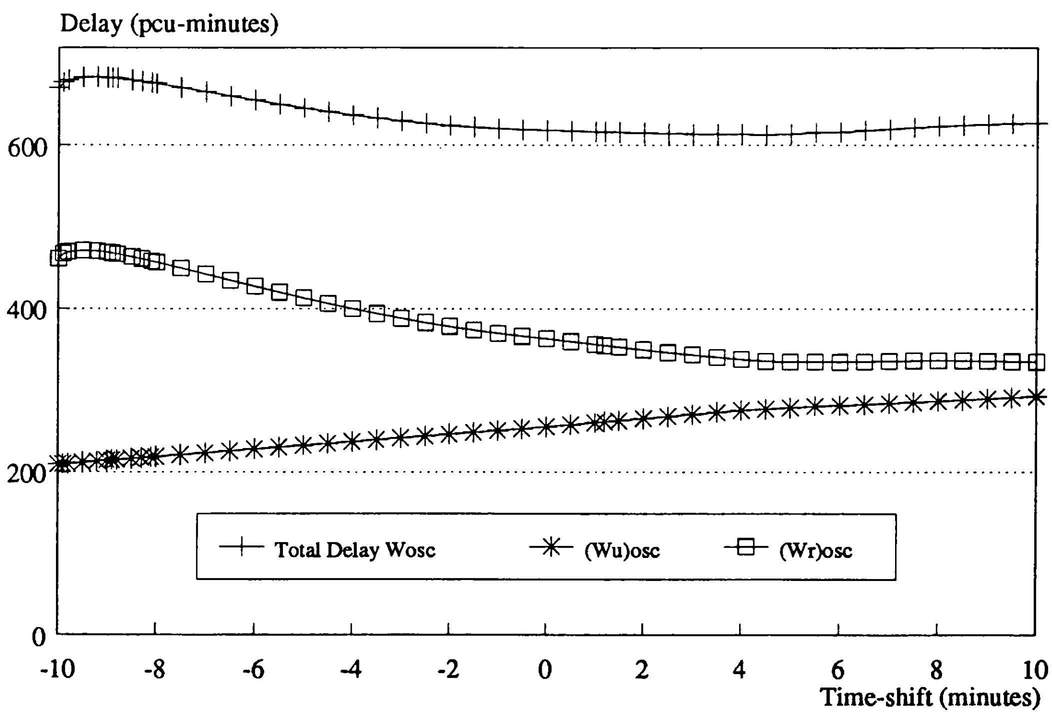


Fig 5.12b Delay Estimated By The OSCADY Delay Formula

These findings suggest that it is difficult to apply analytical methods (methods that require the derivatives $\partial W(\delta t_k)/\partial \delta t_K$ to search the optimal δt_k. The fact that any change in the signal settings or time-shift for one period will result in changes in the delays in the subsequent periods implies that it is also difficult to optimise the elements of $\{\delta t_k\}$ simultaneously, especially when the number of time periods p is large, and the amount of computation required to evaluate any changes in the δt_k is big. Hence it is only practical to optimise $\{\delta t_k\}$ sequentially. An optimisation algorithm for this purpose using the Golden Section method is introduced in §5.6.3.

§5.6.3 The Optimisation Method – Golden Section Method

To optimise $\{\delta t_k\}$ sequentially under fixed $\{\underline{\lambda_k}'\}$, first we consider the optimisation of δt_k, k=1, 2, ,···, p-1. The initial feasible region for δt_k is defined according to the current δt_{k-1}, δt_{k+1} and the constraints (5.3), and will be denoted by the interval (a_k, b_k).

$$\text{Then } a_k = \begin{cases} -T_k & \text{if } \delta t_{k-1} \leq 0 \\ -T_k + \delta t_{k-1} & \text{if } \delta t_{k-1} > 0 \end{cases} \tag{5.10a}$$

$$\text{Then } b_k = \begin{cases} T_{k+1} + \delta t_{k+1} & \text{if } \delta t_{k+1} \leq 0 \\ T_{k+1} & \text{if } \delta t_{k+1} > 0 \end{cases} \tag{5.10b}$$

If $W(\delta t_k)$ has only one turning value, a minimum, in (a_k, b_k), then the optimal δt_k can be found by using the Golden Section method. However, since it is possible that there are other local minimum points in this interval for δt_k, local optimisation in this interval may give a suboptimal δt_k. Based on the assumption that the optimal δt_k should be in the neighbourhood of the current value δt_{k0}, say, the optimal δt_k can be found by iterative reduction of the interval while optimising δt_k using the Golden Section method. Such an algorithm for optimising $\{\delta t_k\}$ can be described as follows:

ALGORITHM 2 *(sequential optimisation of $\{\delta t_k\}$):*

Step 0: Define the initial feasible region (a_k, b_k) for δt_k (k = 1, 2, ···, p-1).

Step 1: Optimise $\{\delta t_k\}$ ($k = 1, 2, \cdots, p-1$).

step 1.1 Choose a small value ε_2 for use in a stopping criterion.

step 1.k: Optimise δt_k ($k = 1, 2, \cdots, p-1$).

step 1.k.0 Evaluate the initial performance index $W_0 = W(\delta t_{k0})$.

step 1.k.1 Let the current feasible region be $(a_k|_f, b_k|_f)$, where f is the number of times that (a_k, b_k) is reduced ($f = 0, 1, \cdots$).

step 1.k.2 Let the current search interval for δt_k be (a_{kh}, b_{kh}), where h is the current iteration number ($h = 1, 2, \cdots$, $a_{k1}=a_k|_f$ and $b_{k1}=b_k|_f$). The length of the interval is $\alpha_h = b_{kh}-a_{kh}$. Then find two trial points ξ_{h1} and ξ_{h2}, which are given by

$$\xi_{h1} = b_{kh} - \tau\alpha_h$$

$$\xi_{h2} = a_{kh} + \tau\alpha_h$$

Where $\tau = \dfrac{\sqrt{5} - 1}{2} \approx 0.618$ is the factor by which the interval is reduced.

step 1.k.3 Compare $W(\xi_{h1})$ and $W(\xi_{h2})$.

If $W(\xi_{h1})<W(\xi_{h2})$, then let $a_{k(h+1)} = a_{kh}$, $b_{k(h+1)} = \xi_{h2}$; hence $\xi_{(h+1)2} = \xi_{h1}$, $\xi_{(h+1)1}=b_{k(h+1)}-\tau\alpha_{(h+1)}$. Take the current solution as $\delta t_{kh} = \xi_{h1}$.

otherwise let $a_{k(h+1)} = \xi_1$, $b_{k(h+1)} = b_{kh}$; hence $\xi_{(h+1)1} = \xi_{h2}$, $\xi_{(h+1)2}=a_{k(h+1)}+\tau\alpha_{(h+1)}$, and take the current solution as $\delta t_{kh} = \xi_{h2}$.

step 1.k.4 If $\{W(\xi_{h1})-W(\xi_{h2})\}/W(\xi_{h1}) > \varepsilon_2$, go to step 1.k.3. Otherwise compare W_0 and $W_h = W(\delta t_{kh})$.

If $W_h > W_0$, reduce the feasible region by:

If $\delta t_{kh} > \delta t_{k0}$, let $a_k|_{f+1} = a_k|_f$, $b_k|_{f+1} = \delta t_{kh}$;
If $\delta t_{kh} < \delta t_{k0}$, let $a_k|_{f+1} = \delta t_{kh}$, $b_k|_{f+1} = b_k|_f$.
Then got to step 1.k.1 to search δt_k again in the revised feasible region.

Otherwise stop and take δt_{kh} as the solution.

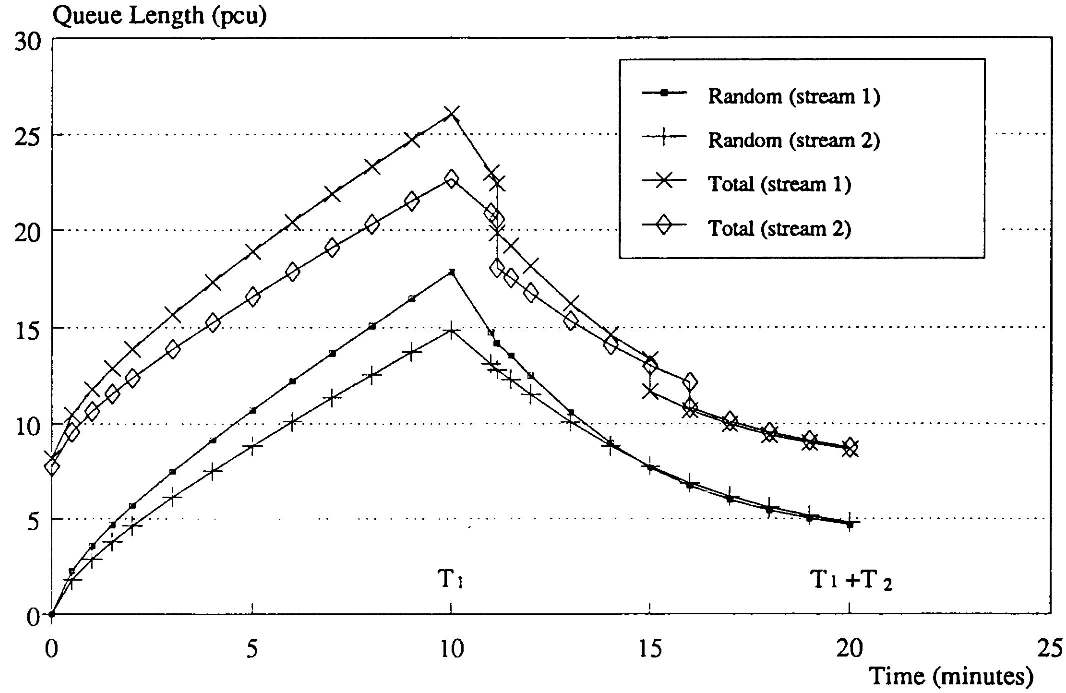
Such an algorithm will generate a sequence of feasible region $(a_k|_0, b_k|_0)$, $(a_k|_1, b_k|_1)$, $(a_k|_2, b_k|_2)$, $\cdots$, which will converge to the neighbourhood of δt_{k0} in which the only turning value is a minimum, and hence a performance index which is no worse than W_0 can be found.

§5.6.4 Results For The Present Example

Using this algorithm for the present example, the resulting W is nearly the same as that given by the numerical calculation, and δt_1 is a little different since the curve $W(\delta t_k)$ is almost horizontal around the optimum. The result can be summarised in Table 5.6, where the percentage reduction in delay is still estimated with respect to the total delay W given by the initial settings $\{\underline{\lambda_k}\}$. By comparing Table 5.6 and Table 5.3 it can be seen that only a very small improvement in W is made by shifting δt_1 in this example. The queues remaining at the end of the second period are

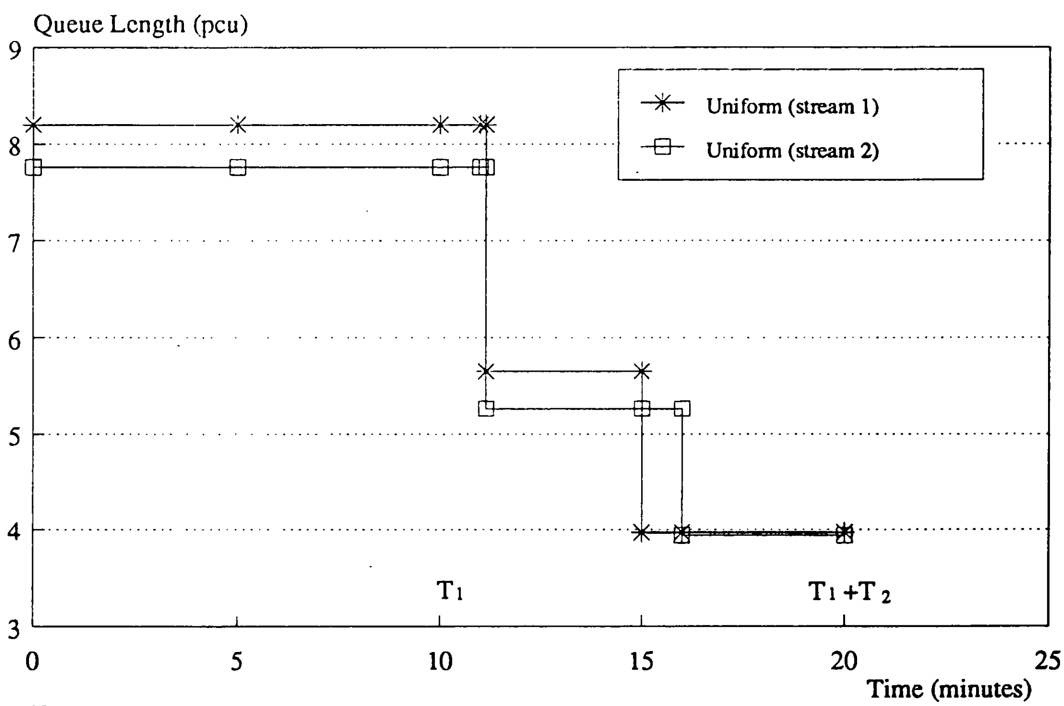
however reduced by about 5 per cent. Fig 5.13 shows the evolution of the queue lengths, from which by careful comparison with Fig 5.10 it can be seen that by making the time-shift the random queues after T_1 are reduced, and the fact that total delay is reduced shows that these reductions in the random queues after T_1 overweigh the increase in uniform queues between T_1 and $T_1+\delta t_1$.

Fig 5.13 Queue Lengths
After Shifting t_1 By 68.74 Seconds



(a) Total and random queue lengths

Fig 5.13 Queue Lengths
After Shifting t_1 By 68.74 Seconds



(b) Uniform queue lengths

Table 5.6 The Signal Settings After Re-optimising $\{\underline{\lambda}_k'\}$ and Shifting δt_1

	Time-shift δt_1=68.74 seconds	
	Period T_1	Period T_2
Cycle Time (Seconds)	c_1 = 120.00	c_2 = 81.78
Allocations of Green Times $\underline{\lambda}_1, \underline{\lambda}_2$	λ_{11}' = 0.5642 λ_{21}' = 0.3691	λ_{12}' = 0.5380 λ_{22}' = 0.3641
Final Random Queue Length	L_{r11} = 17.85 L_{r21} = 14.89	L_{r12} = 4.69 L_{r22} = 4.80
Final Uniform Queue Length	L_{u11} = 8.20 L_{u21} = 7.76	L_{u12} = 3.97 L_{u22} = 3.94
Total Delay	W_1 = 350.96	W_2 = **271.91**
(pcu·minutes)	W_{new} = **622.51** W_{osc} = **616.28**	
Percentage Reduction* In Total Delay $100 \cdot \delta W / W$	**5.64** % by the new formula **5.78** % by the OSCADY formula	

*compared with initial settings (Table 5.2)

§5.7 A SEQUENTIAL OPTIMISATION METHOD

§5.7.1 Introduction

The discussion so far on the optimisation of traffic signal settings for periods of time-varying demand has been confined to optimising $\{\underline{\lambda}_k'\}$ without changing $\{\delta t_k\}$, or optimising $\{\delta t_k\}$ without changing $\{\underline{\lambda}_k'\}$. However, since $\{\underline{\lambda}_k'\}$ and $\{\delta t_k\}$ are interrelated, the signal settings given by either Algorithm 1 or Algorithm 2 are only local optima and can be improved if the two algorithms are combined together. Hence the following sequential optimisation method is proposed.

§5.7.2 The Algorithm

ALGORITHM 3 *(Sequential Optimisation of $\{\underline{\lambda}_k'\}$ and $\{\delta t_k\}$)*

Step 0: Find the initial solution $\{\underline{\lambda}_k'\}=\{\underline{\lambda}_k\}$ and $\{\delta t_k\}=\{0\}$, where $\{\underline{\lambda}_k\}$ is the succession of traffic signal settings that give the optimal performance index for each single time period T_k, k=1, 2, ···, p.

Choose a small value ε_3 for use in a stopping criterion.

Step 1: For current $\{\delta t_k\}$, optimise $(\underline{\lambda_1}', \underline{\lambda_2}', \cdots, \underline{\lambda_p}')=\{\underline{\lambda_k}'\}$ using Algorithm 1.

Step 2: Let the current Performance Index be W_c.

For the current $\{\underline{\lambda_k}'\}$, optimise $\{\delta t_k\}$. Let the resulting Performance Index be W_{new}. If $(W_c - W_{new}) > \varepsilon_3 \cdot W_c$, go to step 1. Otherwise stop.

§5.7.3 Results For The Present Example

The results after implementing Algorithm 3 are summarised in Table 5.7 and the evolution of the queues is shown in Fig 5.14.

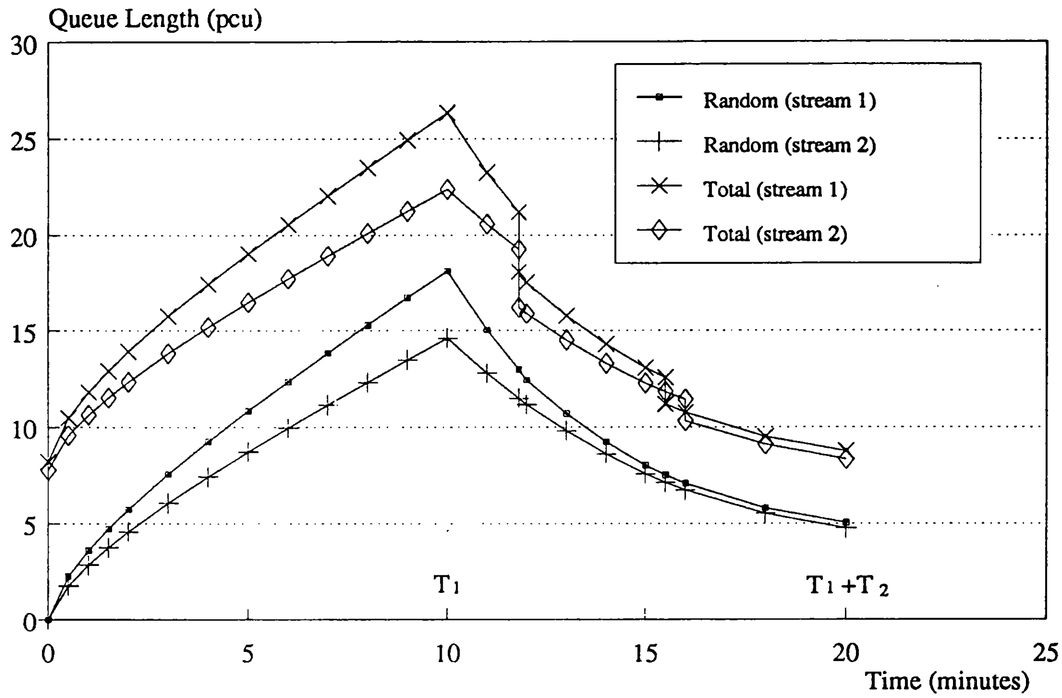
Table 5.7 The Signal Settings After Sequential Optimisation of $\{\underline{\lambda_k}'\}$ & δt_1

	Time-shift $\delta t_1 = 106.3$ seconds	
	Period T_1	Period T_2
Cycle Time (Seconds)	$c_1 = 120.00$	$c_2 = 73.48$
Allocations of Green Times $\underline{\lambda_1}, \underline{\lambda_2}$	$\lambda_{11}' = 0.5631$ $\lambda_{21}' = 0.3702$	$\lambda_{12}' = 0.5298$ $\lambda_{22}' = 0.3614$
Final Random Queue Length	$L_{r11} = 18.13$ $L_{r21} = 14.62$	$L_{r12} = 5.05$ $L_{r22} = 4.76$
Final Uniform Queue Length	$L_{u11} = 8.20$ $L_{u21} = 7.77$	$L_{u12} = 3.69$ $L_{u22} = 3.57$
Total Delay (pcu·minutes)	$W_1 = 351.10$	$W_2 = $ **270.78**
	$W_{new} =$ **621.88** $W_{osc} =$ **615.19**	
Percentage Reduction* In Total Delay $100 \cdot \delta W/W$	**5.80** % by the new formula **5.93** % by the OSCADY formula	

*compared with initial settings (Table 5.2)

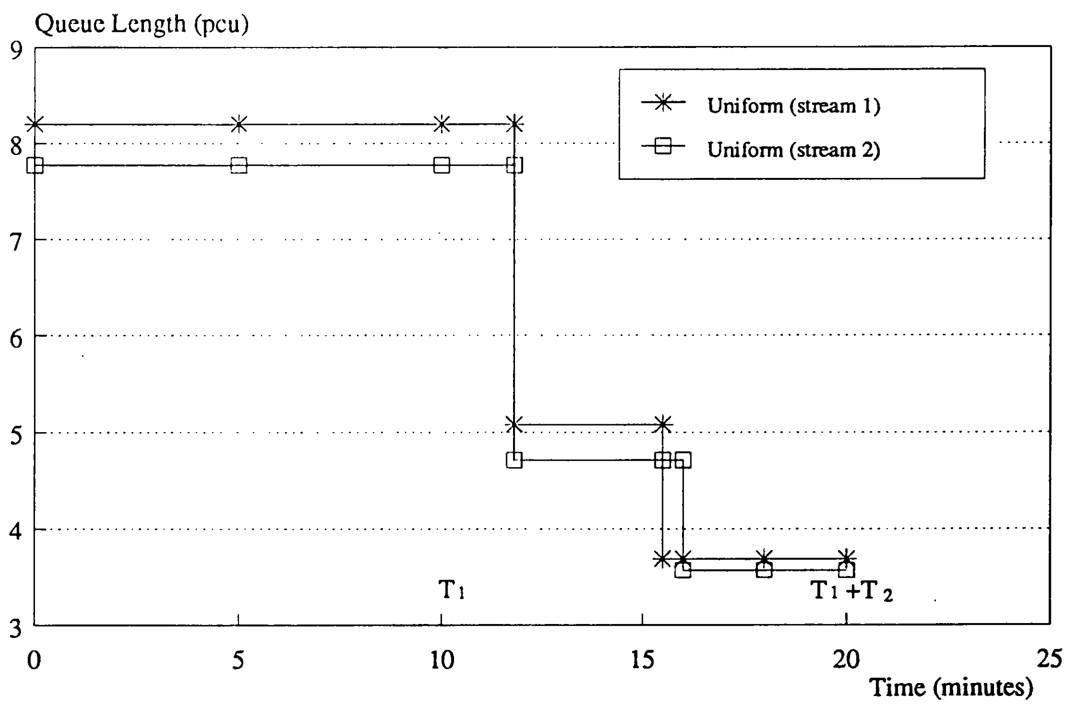
It can be seen from Tables 5.6 and 5.7 that the total delay is reduced very slightly further. Moreover, the queues at the end of T_2 are reduced by a further 6 per cent.

Fig 5.14 Queue Lengths
After Sequential Optimisation



(a) Total and random queue lengths

Fig 5.14 Queue Lengths
After Sequential Optimisation



(b) Uniform queue lengths

§5.8 TAKING INTO ACCOUNT THE SUBSEQUENT CONDITIONS

§5.8.1 Introduction

Although in mathematical terms, the traffic signal settings can be optimised with respect to total delay for periods of time-varying demand for any given flow pattern and junction layout by using Algorithm 3, some practical factors must also be considered in the optimisation process, one of which is the queues left at the end of the last period T_p. These queues will influence the delays in the period after T_p, if they are not small enough to be neglected. However, Algorithm 3 takes no account of the final queues and the delays in the subsequent periods after T_p, and it is possible that unnecessarily longer final queues will occur as a result of such an algorithm, especially when the degree of saturation in T_p is high, which will cause bigger delays than necessary after T_p. Hence it is necessary to make further improvement to Algorithm 3 so that, whilst keeping the total delay W as the performance index, some consideration can also be given to the delays in the subsequent periods after T_p.

In the calculation of signal settings for a succession of periods in practice, two requirements of the flow data are that the data should be specified for as many time periods as is practicable, and the flows in the last period should be low enough for the junction to be undersaturated in T_p, and this period should be long enough for equilibrium to be approached during it. This is so that there will not be big delays in the subsequent periods that could be influenced by the choice of timings for periods T_1 to T_p. Hence a good knowledge about the arrival pattern is preferable and it is reasonable to assume that *the last period is always undersaturated when setting traffic signals for periods of time-varying demand.*

Under this assumption, the queue length for each stream will approach an equilibrium value in T_p if T_p is long enough. The delay-minimising algorithm can then be applied to a series of time periods at the end of which all the streams are in equilibrium. The resulting signal settings will be the global optimal solution which can give the minimum estimated total delay over such a series of time periods.

However, in practice, the engineer may have difficulty in specifying the length of the last period T_p so that the queues in each stream are in equilibrium at the end of T_p, or may give a shorter T_p than is required for this. In the latter case it may be reasonable to *assume that the flow levels for all the streams are the same after T_p as they were during T_p.*

Hence when optimising the signal settings for p time periods, if it happens that at the end of T_p, some streams still have longer queues than their equilibrium queue lengths, then T_p can be extended to, say, T_{pe}, and the signal settings re-optimised for $T_1, T_2, \cdots, T_{p-1}, T_{pe}$. This process can be repeated until all the queues are in equilibrium before or at the end of T_{pe}. In this way some account is taken of the subsequent delays.

To sum up, the following two assumptions are made to find the signal settings that can take account of queues remaining at the end of T_p.

(1) In the last period T_p the junction is undersaturated;

(2) The flows in T_p may last longer than T_p.

Based on these two assumptions, an improved algorithm is now proposed.

§5.8.2 A Further Algorithm

To accommodate the delays in the subsequent periods after T_p, the following algorithm can be implemented.

ALGORITHM 4 *(Sequential Optimisation of $\{\underline{\lambda}_k'\}$ and $\{\delta t_k\}$ With Allowance For Subsequent Delays)*

Step 0: Let $T_{pe} = T_p$. Choose a small number ε_4 (e.g. $\varepsilon_4 = 0.2$) for use in a stopping criterion.

Step 1: Implement Algorithm 3 to find the solution $\{\underline{\lambda}_k'\}$ and $\{\delta t_k\}$ for the time periods $T_1, T_2, \cdots, T_{pe}$.

Step 2: Calculate the final random queue length L_{rjp} at the end of T_{pe} for each stream j (j = 1, 2, ⋯, n), and the equilibrium random queue length L_{rejp}.

Step 3: If for all j (j = 1, 2, ⋯, n), $(L_{rjp}-L_{rejp})<\varepsilon_4 \cdot L_{rejp}$, then the current signal settings are the final solution and go to step 4. Otherwise extend T_{pe} by

Step 3.1: Let $T_{pe} = T_{pe} + \delta t_{pe}$, where δt_{pe} is an increment step (e.g. δt_{pe} = 5 minutes).

Step 3.2: same as step 2.

Step 3.3: If $(L_{rjp}-L_{rcjp}) < \varepsilon_4 \cdot L_{rcjp}$ holds for all j (j = 1, 2, ⋯, n), then go to step 1; otherwise go to step 3.1.

Step 4: evaluate W for periods $T_1, T_2, \cdots, T_p$ and stop.

§5.8.3 Results For The Present Example

As shown in Table 5.7, by Algorithm 3, at the end of T_2, the random queue length for stream 1 and 2 are:

$$L_{r12} = 5.05 > L_{rc12} = 2.87$$

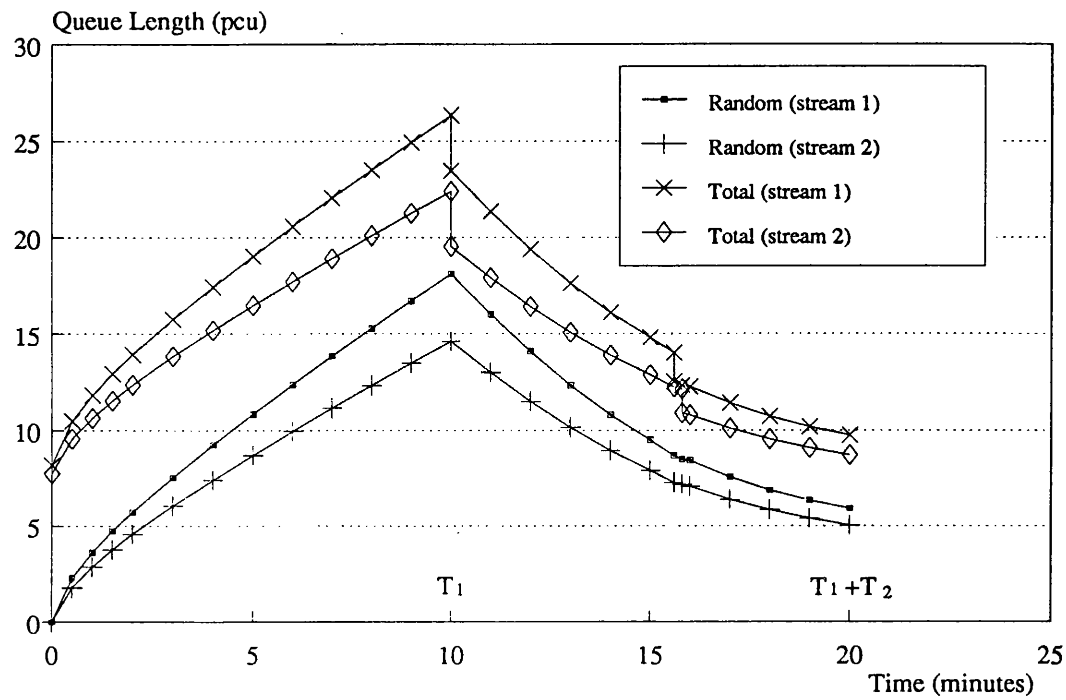
$$L_{r22} = 4.76 > L_{rc22} = 2.44$$

Hence by the end of T_2, the queues for both streams are higher than their equilibrium values, and the signal settings should be modified so that the delays after T_2 can also be taken into account in the optimisation procedure.

When implementing Algorithm 4, in the optimisation process, T_2 is extended to T_{2c} = 46 minutes, at the end of which the queues are approaching (with a maximum error ε_4 = 0.2) equilibrium. The signal settings optimised for periods (T_1+ T_{2c}) are summarised in Table 5.8, where the total delay W_{new} and W_{osc} are evaluated for the periods T_1, T_2. The evolution of the queues is given in Fig 5.15.

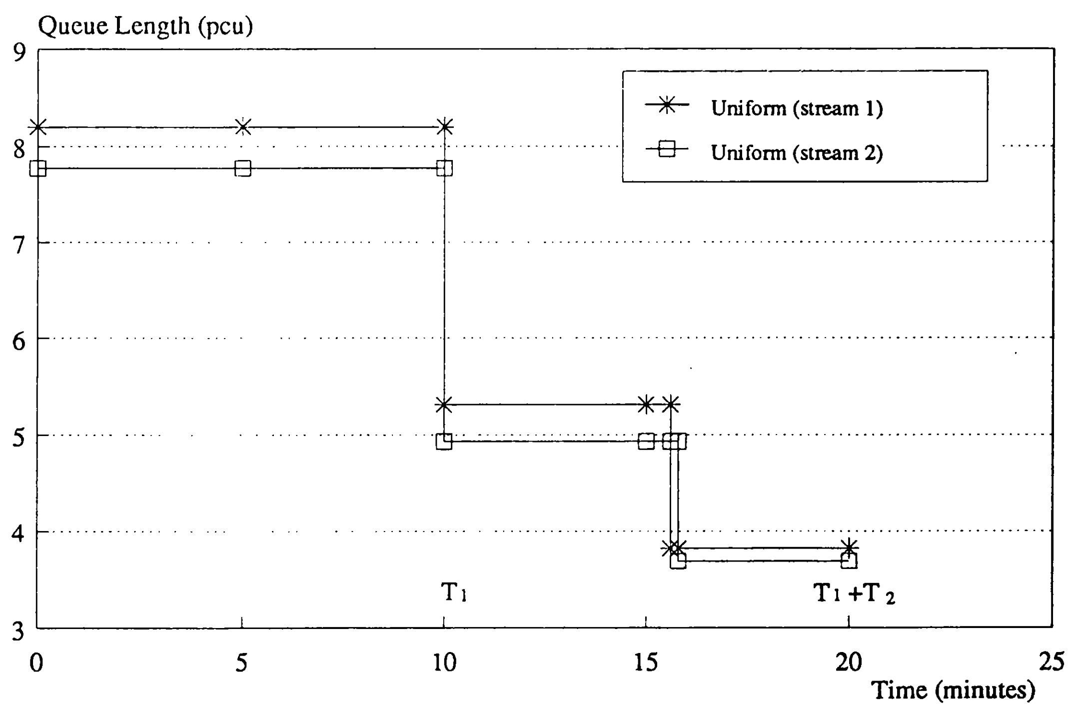
It can be seen from Table 5.8 that the total delay is increased compared with Table 5.7 as a result of accounting for the subsequent delays. The

Fig 5.15 Queue Lengths Over T1 And T2
When Optimisation Is Over 56 Minutes



(a) Total and random queue lengths

Fig 5.15 Queue Lengths Over T1 And T2
When Optimisation Is Over 56 Minutes



(b) Uniform queue lengths

time-shift is zero which means it is now unnecessary to shift the change in signal settings in this example. Because the cycle time in the second period is longer than before, the capacity provided is higher, as a result, the equilibrium queues become:

$L_{re12} = 2.80 < 2.87$

$L_{re22} = 2.32 < 2.44$

i.e. the signal settings given by Algorithm 4 will lead to shorter equilibrium queue lengths than those given by Algorithm 3, hence the subsequent delays are smaller. The resulting delays are minimum for the periods T_1+T_{2c}= 56 minutes rather than for only $T_1 + T_2$ = 20 minutes.

Table 5.8 The Signal Settings After Sequential Optimisation of $\{\underline{\lambda_k}'\}$ & δt_1 Taking Into Account The Subsequent Delays

	Time-shift δt_1=0.0 seconds	
	Period T_1	Period T_2
Cycle Time (Seconds)	c_1 = 120.00	c_2 = 76.73
Allocations of Green Times $\underline{\lambda_1}, \underline{\lambda_2}$	λ_{11}'= 0.5631 λ_{21}'= 0.3702	λ_{12}'= 0.5317 λ_{22}'= 0.3640
Final Random Queue Length	L_{r11} = 18.14 L_{r21} = 14.62	L_{r12} = 5.93 L_{r22} = 5.04
Final Uniform Queue Length	L_{u11} = 8.20 L_{u21} = 7.77	L_{u12} = 3.82 L_{u22} = 3.69
Total Delay (pcu·minutes)	W_1 = **351.06**	W_2 = **273.68**
	W_{new} = **624.74** W_{osc} = **620.31**	
Percentage Reduction* In Total Delay $100 \cdot \delta W / W$	**5.36** % by the new formula **5.15** % by the OSCADY formula	

***compared with initial settings (Table 5.2)**

Hence when the subsequent delays after T_p need to be considered, the resulting signal settings may give higher delays and longer final queues for the periods T_1, T_2, ···, T_p than when such delays are not considered.

§5.9 MORE EXAMPLE CALCULATIONS FOR TWO-PERIOD CASES

§5.9.1 Introduction

To demonstrate further the optimisation method described above, some more example calculations are made for typical two-period cases. The traffic pattern is that the junction is oversaturated in the first period T_1, and returns to undersaturation in the second period T_2. Apart from $\underline{\lambda_1}'$ and $\underline{\lambda_2}'$, only one value of time-shift needs to be decided, which is δt_1 (since $\delta t_0 = 0$). The initial settings are $\underline{\lambda_1}'=\underline{\lambda_1}$, $\underline{\lambda_2}'=\underline{\lambda_2}$, and $\delta t_1=0$.

$\underline{\lambda_1}$ and $\underline{\lambda_2}$ are obtained by Algorithm 0, through optimising the total delay in T_1 and T_2 separately. Algorithm 4 introduced in §5.8 is used for the sequential optimisation of $\underline{\lambda_1}'$, $\underline{\lambda_2}'$ and δt_1, but to reduce the period over which calculation is needed, the error limit ε_4 is set to 0.2, and no extension is made if the final random queues are shorter than 120% of their equilibrium values. However, if an extension for T_2 is needed, a maximum value for T_{2e} is imposed, which is $(T_{2e})_{max}$ = 50 minutes. This constraint is introduced to make sure that the sum of T_1 and T_{2e} is not going to exceed one hour, since it is not likely that the traffic flows will stay unchanged for more than one hour. After the optimisation, the queues and delays are estimated by both the OSCADY formulae and the new delay expression. If an extension for T_2 is made, then the total delay are evaluated over (T_1+T_{2e}) as well as over T_1+T_2.

The three example junctions introduced in Chapter 4 will again be used here for example calculations. All the parameters, such as the maximum cycle time, the minimum green times for each stage and the lost times after each stage, etc, are kept the same here.

In each case for each junction, calculations are also made for the program OSCADY and the results are compared by using the criterion $\delta W/W$.

§5.9.2 Junction 1 — Symmetrical Crossroads

The additional data are given in Table 5.9, and the results are given in Table 5.10.

Table 5.9 Additional Data For Example Junction 1

	Period T_1	Period T_2
Length Of Time Period	T_1=10 minutes	T_2=10 minutes
Flow Ratio In Stream 1	$y_{11} = 0.6$	$y_{12} = 0.3$
Flow Ratio In Stream 2	$y_{21} = 0.4$	$y_{22} = 0.2$
Initial Queue Length In Stream 1	$L_{r10} = 0.0$	
Initial Queue Length In Stream 2	$L_{r20} = 0.0$	

After the optimisation, δt_1=0.0 seconds. It is clear from Table 5.10 that the optimisation method offers about 1% reduction in the total delay compared with the case when no sequential optimisation was made, and about 3% reduction in total delay compared with the OSCADY method, which gives very different cycle times because it is maximising capacity in the first period and minimising equilibrium delay in the second. However, at the end of T_2, the queues are higher than 120% of their equilibrium values (with $L_{r12} = 1.15 > L_{re12} = 0.6$, $L_{r22} = 0.94 > L_{re22} = 0.43$), hence an extension is made for T_2 to the maximum value, i.e. $T_{2e} = 50$ minutes. The signal settings and the relevant results for the period ($T_1 + T_{2e}$) are listed in Table 5.11.

Table 5.10 Results for Example Junction 1 Optimised Over T_1 and T_2 Only

δt_1 = 0.0 seconds; $T_{2e} = T_2$			PRESENT METHOD		OSCADY
			Initial Settings	Optimised Settings	
PERIOD T_1	Cycle Time (Seconds)		87.49	106.34	120.00
PERIOD T_1	Proportion of Cycle Effectively Green For Stage	1	0.5583	0.5616	0.5600
		2	0.3502	0.3632	0.3733
PERIOD T_1	Proportion of Cycle Effectively Green For Stream	1	0.5583	0.5616	0.5600
		2	0.3502	0.3632	0.3733
PERIOD T_1	Final Uniform Queue Length For Stream (pcu)	1	5.99	7.27	8.21
		2	5.53	6.83	7.80
PERIOD T_1	Final Random Queue Length For Stream (pcu)	1	19.33	18.53	18.91
		2	19.91	16.41	13.86
PERIOD T_2	Cycle Time (Seconds)		53.13	50.27	36.10
PERIOD T_2	Proportion of Cycle Effectively Green For Stage	1	0.4793	0.4853	0.4598
		2	0.3701	0.3555	0.3186
PERIOD T_2	Proportion of Cycle Effectively Green For Stream	1	0.4793	0.4853	0.4598
		2	0.3701	0.3555	0.3186
PERIOD T_2	Final Uniform Queue Length For Stream (pcu)	1	1.71	1.59	1.25
		2	1.46	1.45	1.16
PERIOD T_2	Final Random Queue Length For Stream (pcu)	1	1.26	1.15	1.50
		2	0.95	0.94	1.31
Total Delay (Pcu·Minutes) Estimated By	New Formula		**444.85**	**437.79**	**448.60**
	OSCADY Formula		**434.22**	**430.03**	**445.01**
Percentage Reduction In Total Delay Estimated By	New Formula			**1.59%**	**2.41%**
	OSCADY Formula			**0.96%**	**3.37%**

Table 5.11 Results for Example Junction 1 Optimised After Extending T_2 To 50 minutes

δt_1 = -43.0 seconds* $T_{2e} = T_2$			PRESENT METHOD Initial Settings	PRESENT METHOD Optimised Settings	OSCADY
PERIOD T_1	Cycle Time (Seconds)		87.49	100.04	120.00
	Proportion of Cycle Effectively Green For Stage	1	0.5583	0.5590	0.5600
		2	0.3502	0.3611	0.3733
	Proportion of Cycle Effectively Green For Stream	1	0.5583	0.5590	0.5600
		2	0.3502	0.3611	0.3733
	Final Uniform Queue Length For Stream (pcu)	1	5.99	2.72	8.21
		2	5.53	2.41	7.80
	Final Random Queue Length For Stream (pcu)	1	19.33	21.53	18.91
		2	19.91	17.91	13.86
PERIOD T_{2e}	Cycle Time (Seconds)		38.81	39.33	36.10
	Proportion of Cycle Effectively Green For Stage	1	0.4630	0.4679	0.4598
		2	0.3308	0.3287	0.3186
	Proportion of Cycle Effectively Green For Stream	1	0.4630	0.4679	0.4598
		2	0.3308	0.3287	0.3186
	Final Uniform Queue Length For Stream (pcu)	1	1.33	1.33	1.25
		2	1.21	1.23	1.16
	Final Random Queue Length For Stream (pcu)	1	0.83	0.81	0.85
		2	0.68	0.68	0.73
Total Delay (Pcu·Minutes) Estimated By	New Formula		**573.52**	**570.58**	**578.74**
	OSCADY Formula		**566.74**	**563.96**	**575.54**
Percentage Reduction In Total Delay Estimated By	New Formula			**0.51%**	**1.40%**
	OSCADY Formula			**0.49%**	**2.00%**

*Note: The signal timings given in the table apply in the periods T_1' and T_2' but the final queue lengths for the first period are those at the end of T_1.

It can be seen from Table 5.11 that after extending T_2, the signal settings obtained for the second period before and after the optimisation and those given by OSCADY are nearly the same and thus they give very similar results. This is because in the second period the junction has nearly reached its equilibrium condition, and the signal settings are nearly the same as the equilibrium settings as given by OSCADY. As T_2 becomes longer and the signal settings calculated for it change, a negative time-shift is obtained, which means that there is some small benefit if $\underline{\lambda_1}'$ are replaced by $\underline{\lambda_2}'$ earlier than the change of flows.

§5.9.3 Junction 2 — Asymmetrical Crossroads

(1) case 1: the wider road has a higher y value.

The additional data are given in Table 5.12, and the results are given in Table 5.13.

Table 5.12 Additional Data For Example Junction 2

	Period T_1	Period T_2
Length Of Time Period	T_1=10 minutes	T_2=10 minutes
Flow Ratio In Stream 1	$y_{11} = 0.6$	$y_{12} = 0.3$
Flow Ratio In Stream 2	$y_{21} = 0.4$	$y_{22} = 0.2$
Initial Queue Length In Stream 1	$L_{r10} = 0.0$	
Initial Queue Length In Stream 2	$L_{r20} = 0.0$	

Table 5.13 Results for Example Junction 2 Optimised Over T_1 and T_2 Only

δt_1 = 0.0 seconds $T_{2e} = T_2$			PRESENT METHOD Initial Settings	PRESENT METHOD Optimised Settings	OSCADY
PERIOD T_1	Cycle Time (Seconds)		81.48	101.26	120.00
	Proportion of Cycle Effectively Green For Stage	1	0.6243	0.6063	0.5600
		2	0.2775	0.3147	0.3733
	Proportion of Cycle Effectively Green For Stream	1	0.6243	0.6063	0.5600
		2	0.2775	0.3147	0.3733
	Final Uniform Queue Length For Stream (pcu)	1	9.58	13.08	16.43
		2	4.54	6.06	7.80
	Final Random Queue Length For Stream (pcu)	1	8.82	12.96	33.18
		2	42.12	30.43	13.86
PERIOD T_2	Cycle Time (Seconds)		54.62	47.65	33.60
	Proportion of Cycle Effectively Green For Stage	1	0.3738	0.4052	0.4643
		2	0.4797	0.4270	0.2976
	Proportion of Cycle Effectively Green For Stream	1	0.3738	0.4052	0.4643
		2	0.4797	0.4270	0.2976
	Final Uniform Queue Length For Stream (pcu)	1	5.10	4.01	2.99
		2	1.03	1.09	1.47
	Final Random Queue Length For Stream (pcu)	1	2.40	1.68	1.34
		2	0.84	0.85	1.83
Total Delay (Pcu·Minutes) Estimated By	New Formula		**592.30**	**569.80**	**629.48**
	OSCADY Formula		**576.73**	**561.65**	**624.08**
Percentage Reduction In Total Delay Estimated By	New Formula			**3.80%**	**9.47%**
	OSCADY Formula			**3.61%**	**9.99%**

Similar to example junction 1, by the end of T_2, the queues are still longer than 120% of their equilibrium values, hence T_2 is extended to the maximum limit of 50 minutes. After the extension, the optimisation results for the periods T_1 and T_{2e} are given in Table 5.14 below.

Table 5.14 Results for Example Junction 2 Optimised After Extending T_2 To 50 minutes

δt_1 = -24.0 seconds* T_{2e} = T_2			PRESENT METHOD Initial Settings	PRESENT METHOD Optimised Settings	OSCADY
PERIOD T_1	Cycle Time (Seconds)		81.48	93.29	120.00
	Proportion of Cycle Effectively Green For Stage	1	0.6243	0.6112	0.5600
		2	0.2775	0.3030	0.3733
	Proportion of Cycle Effectively Green For Stream	1	0.6243	0.6112	0.5600
		2	0.2775	0.3030	0.3733
	Final Uniform Queue Length For Stream (pcu)	1	9.58	4.95	16.43
		2	4.54	2.23	7.80
	Final Random Queue Length For Stream (pcu)	1	8.82	15.97	33.18
		2	42.12	33.72	13.86
PERIOD T_{2e}	Cycle Time (Seconds)		36.19	36.09	33.60
	Proportion of Cycle Effectively Green For Stage	1	0.4356	0.4454	0.4643
		2	0.3433	0.3328	0.2976
	Proportion of Cycle Effectively Green For Stream	1	0.4356	0.4454	0.4643
		2	0.3433	0.3328	0.2976
	Final Uniform Queue Length For Stream (pcu)	1	2.74	2.64	2.30
		2	1.08	1.12	1.15
	Final Random Queue Length For Stream (pcu)	1	0.95	0.89	0.80
		2	0.75	0.77	0.97
Total Delay (Pcu·Minutes) Estimated By	New Formula		**749.46**	**742.86**	**804.64**
	OSCADY Formula		**738.34**	**734.65**	**799.73**
Percentage Reduction In Total Delay Estimated By	New Formula			**0.88%**	**7.68%**
	OSCADY Formula			**0.50%**	**8.13%**

*Note: The signal timings given in the table apply in the periods T_1' and T_2' but the final queue lengths for the first period are those at the end of T_1.

It can be seen that even though after the extension, the difference between the results before and after the sequential optimisation is small, the optimised signal settings still give better results than OSCADY. The general pattern of differences between the signal settings given by the new method and by OSCADY and the resulting differences in queue-lengths in the two streams are as discussed in Chapter 4.

(2) case 2: The narrower road has a higher y value.

The additional data are given in Table 5.15.

Table 5.15 Additional Data For Example Junction 2

	Period T_1	Period T_2
Length Of Time Period	T_1=10 minutes	T_2=10 minutes
Flow Ratio In Stream 1	$y_{11} = 0.4$	$y_{12} = 0.2$
Flow Ratio In Stream 2	$y_{21} = 0.6$	$y_{22} = 0.3$
Initial Queue Length In Stream 1	$L_{r10} = 0.0$	
Initial Queue Length In Stream 2	$L_{r20} = 0.0$	

The results before extending T_2 are given below in Table 5.16.

Table 5.16 Results for Example Junction 2 Optimised Over T_1 and T_2 Only

δt_1 = 0.0 seconds; T_{2e} = T_2			PRESENT METHOD: Initial Settings	PRESENT METHOD: Optimised Settings	OSCADY
PERIOD T_1	Cycle Time (Seconds)		77.66	96.29	120.00
	Proportion of Cycle Effectively Green For Stage	1	0.4107	0.4031	0.3733
		2	0.4863	0.5139	0.5600
	Proportion of Cycle Effectively Green For Stream	1	0.4107	0.4031	0.3733
		2	0.4863	0.5139	0.5600
	Final Uniform Queue Length For Stream (pcu)	1	9.99	12.71	15.60
		2	5.39	6.68	8.21
	Final Random Queue Length For Stream (pcu)	1	9.02	11.12	23.76
		2	40.24	31.82	18.91
PERIOD T_2	Cycle Time (Seconds)		54.99	48.94	32.50
	Proportion of Cycle Effectively Green For Stage	1	0.2679	0.2855	0.3200
		2	0.5866	0.5511	0.4338
	Proportion of Cycle Effectively Green For Stream	1	0.2679	0.2855	0.3200
		2	0.5866	0.5511	0.4338
	Final Uniform Queue Length For Stream (pcu)	1	4.09	3.47	2.09
		2	1.12	1.17	1.24
	Final Random Queue Length For Stream (pcu)	1	1.75	1.37	1.19
		2	1.06	1.06	2.04
Total Delay (Pcu·Minutes) Estimated By	New Formula		**573.66**	**559.76**	**603.66**
	OSCADY Formula		**561.57**	**552.46**	**598.83**
Percentage Reduction In Total Delay Estimated By	New Formula			**2.42%**	**7.28%**
	OSCADY Formula			**1.62%**	**7.74%**

In this case T_2 also needs to be extended so that the final queues can nearly reach equilibrium. The results are given in Table 5.17.

Table 5.17 Results for Example Junction 2 Optimised After Extending T_2 To 50 minutes

δt_1 = -29.7 seconds* $T_{2e} = T_2$			PRESENT METHOD Initial Settings	PRESENT METHOD Optimised Settings	OSCADY
PERIOD T_1	Cycle Time (Seconds)		77.66	88.71	120.00
	Proportion of Cycle Effectively Green For Stage	1	0.4107	0.4057	0.3733
		2	0.4863	0.5041	0.5600
	Proportion of Cycle Effectively Green For Stream	1	0.4107	0.4057	0.3733
		2	0.4863	0.5041	0.5600
	Final Uniform Queue Length For Stream (pcu)	1	9.99	4.24	15.60
		2	5.39	2.46	8.21
	Final Random Queue Length For Stream (pcu)	1	9.02	13.45	23.78
		2	40.24	35.53	18.91
PERIOD T_{2e}	Cycle Time (Seconds)		35.68	35.64	32.50
	Proportion of Cycle Effectively Green For Stage	1	0.3049	0.3103	0.3200
		2	0.4709	0.4652	0.4338
	Proportion of Cycle Effectively Green For Stream	1	0.3049	0.3103	0.3200
		2	0.4709	0.4652	0.4338
	Final Uniform Queue Length For Stream (pcu)	1	2.39	2.35	2.09
		2	1.19	1.21	1.24
	Final Random Queue Length For Stream (pcu)	1	0.79	0.75	0.71
		2	0.91	0.92	1.08
Total Delay (Pcu·Minutes) Estimated By	New Formula		**733.78**	**729.84**	**772.59**
	OSCADY Formula		**726.20**	**723.17**	**767.93**
Percentage Reduction In Total Delay Estimated By	New Formula			**0.54%**	**5.53%**
	OSCADY Formula			**0.42%**	**5.82%**

*Note: The signal timings given in the table apply in the periods T_1' and T_2' but the final queue lengths for the first period are those at the end of T_1.

It can be seen that the pattern of results is quite similar to that in Tables 5.13 and 5.14. In each case, the new method favours traffic in the wider road in the oversaturated period.

§5.9.4 Junction 3 — Chapel Hill Junction

In period 1, the arrival rate for each stream is the same as given by Table 4.14, and in period 2, it is the same as given by Table 4.12. Other parameters are the same as given by Table 4.10 and 4.11. For T_1=10 minutes, T_2=10 minutes, the results are given by Table 5.18a and 5.18b below.

Table 5.18a Results for Example Junction 3 Optimised Over T_1 and T_2 Only – Period 1

δt_1 = 0.0 seconds, T_{2e} = T_2			PRESENT METHOD Initial Settings	PRESENT METHOD Optimised Settings	OSCADY
PERIOD T_1	Cycle Time (Seconds)		72.29	100.48	120.00
	Proportion of Cycle Effectively Green For Stage	1	0.3449	0.3403	0.3225
		2	0.2583	0.2756	0.2450
		3	0.0856	0.1601	0.2450
		4	0.0830	0.0597	0.0500
	Proportion of Cycle Effectively Green For Stream	1	0.1798	0.1294	0.1083
		2	0.0856	0.1601	0.2450
		3	0.8245	0.7752	0.7008
		4	0.3482	0.3403	0.2992
		5	0.4971	0.4500	0.4142
		6	0.8245	0.7752	0.7008
		7	0.3826	0.4636	0.5133
		8	0.3518	0.3454	0.3267
		9	0.0856	0.1601	0.2450
	Final Uniform Queue Length For Stream (pcu)	1	2.76	4.33	5.42
		2	3.14	7.50	11.04
		3	0.46	1.05	2.22
		4	7.96	10.95	12.21
		5	7.48	10.29	12.04
		6	1.21	2.76	5.85
		7	5.45	5.72	5.62
		8	6.37	9.02	10.87
		9	0.01	0.01	0.01
	Final Random Queue Length For Stream (pcu)	1	0.38	1.28	3.14
		2	93.34	44.74	4.42
		3	0.21	0.24	0.33
		4	49.64	53.99	77.12
		5	25.50	46.22	63.02
		6	2.20	13.32	8.01
		7	1.55	0.68	0.48
		8	6.48	7.64	12.38
		9	0.00	0.00	0.00

Table 5.18b Results for Example Junction 3 Optimised
Over T_1 and T_2 Only – Period 2

δt_1 = 0.0 seconds; T_{2e} = T_2			PRESENT METHOD Initial Settings	PRESENT METHOD Optimised Settings	OSCADY
PERIOD T_2	Cycle Time (Seconds)		73.94	65.59	60.10
	Proportion of Cycle Effectively Green For Stage	1	0.2366	0.2718	0.2962
		2	0.1670	0.2157	0.2246
		3	0.2920	0.1694	0.1048
		4	0.0811	0.0915	0.0998
	Proportion of Cycle Effectively Green For Stream	1	0.1758	0.1982	0.2163
		2	0.2920	0.1694	0.1048
		3	0.6201	0.7315	0.7870
		4	0.2550	0.3148	0.3328
		5	0.3854	0.4395	0.4792
		6	0.6201	0.7315	0.7870
		7	0.4969	0.4278	0.3760
		8	0.2434	0.2795	0.3045
		9	0.2920	0.1694	0.1048
	Final Uniform Queue Length For Stream (pcu)	1	0.89	0.74	0.65
		2	1.48	1.75	1.86
		3	0.41	0.18	0.10
		4	6.82	3.31	2.87
		5	4.78	3.52	2.79
		6	1.76	0.78	0.45
		7	3.32	3.81	4.15
		8	4.94	3.98	3.39
		9	0.01	0.01	0.01
	Final Random Queue Length For Stream (pcu)	1	0.03	0.02	0.02
		2	0.97	1.51	0.83
		3	0.02	0.01	0.01
		4	12.97	2.99	5.73
		5	3.61	3.24	3.22
		6	0.69	0.41	0.36
		7	0.42	0.70	1.17
		8	5.78	2.40	1.86
		9	0.00	0.00	0.00
Total Delay (Pcu·Minutes) Estimated By	New Formula		2347.11	2197.53	2482.26
	OSCADY Formula		2271.42	2113.41	2406.55
Percentage Reduction In Total Delay Estimated By	New Formula			6.37%	11.47%
	OSCADY Formula			6.96%	12.18%

It can be seen from Table 5.18 that for this complex situation, a greater improvement is obtained by sequential optimisation of the signal settings even though no time-shift results. The cycle time in T_1 is increased (from

72.3 seconds to 100.5 seconds); but the cycle time in T_2 is decreased (from 73.9 seconds to 65.6 seconds). Although the uniform delay in T_1 is increased, the total random queue length at the end of T_1 is decreased (from 179.3 pcu to 168.1 pcu), which therefore causes less delay in T_2. However, since at the end of T_2, the queue lengths for some streams are still greater than 120% of their equilibrium values, the second period T_2 is extended to T_{2e}=50 minutes. The results are given in Table 5.19a and Table 5.19b.

Table 5.19a Results for Example Junction 3 Optimised
After Extending T_2 To 50 minutes – Period 1

δt_1 = 0.0 seconds, T_{2e} = T_2			PRESENT METHOD		OSCADY
			Initial Settings	Optimised Settings	
PERIOD T_1	Cycle Time (Seconds)		72.29	82.37	120.00
	Proportion of Cycle Effectively Green For Stage	1	0.3449	0.3438	0.3225
		2	0.2583	0.2772	0.2450
		3	0.0856	0.1558	0.2450
		4	0.0830	0.0728	0.0500
	Proportion of Cycle Effectively Green For Stream	1	0.1798	0.1578	0.1083
		2	0.0856	0.1558	0.2450
		3	0.8245	0.7653	0.7008
		4	0.3482	0.3062	0.2992
		5	0.4971	0.4774	0.4142
		6	0.8245	0.7653	0.7008
		7	0.3826	0.4170	0.5133
		8	0.3518	0.3500	0.3267
		9	0.0856	0.1558	0.2450
	Final Uniform Queue Length For Stream (pcu)	1	2.76	3.32	5.42
		2	3.14	6.01	11.04
		3	0.46	0.94	2.22
		4	7.96	8.49	12.21
		5	7.48	8.50	12.04
		6	1.21	2.47	5.85
		7	5.45	5.54	5.62
		8	6.37	7.30	10.87
		9	0.01	0.01	0.01
	Final Random Queue Length For Stream (pcu)	1	0.38	0.59	3.14
		2	93.34	47.50	4.42
		3	0.21	0.25	0.33
		4	49.64	73.16	77.12
		5	25.50	33.80	63.02
		6	2.20	3.66	8.01
		7	1.55	1.04	0.48
		8	6.48	6.80	12.38
		9	0.00	0.00	0.00

Table 5.19b Results for Example Junction 3 Optimised
After Extending T_2 To 50 minutes – Period 2

δt_1 = 0.0 seconds, T_{2e} = T_2			PRESENT METHOD Initial Settings	PRESENT METHOD Optimised Settings	OSCADY
PERIOD T_{2e}	Cycle Time (Seconds)		54.49	54.04	60.10
	Proportion of Cycle Effectively Green For Stage	1	0.2692	0.2718	0.2962
		2	0.1581	0.1827	0.2246
		3	0.1598	0.1292	0.1048
		4	0.1101	0.1110	0.0998
	Proportion of Cycle Effectively Green For Stream	1	0.2386	0.2406	0.2163
		2	0.1598	0.1292	0.1048
		3	0.7209	0.7505	0.7870
		4	0.2774	0.3029	0.3328
		5	0.4711	0.4753	0.4792
		6	0.7209	0.7505	0.7870
		7	0.3693	0.3637	0.3760
		8	0.2784	0.2810	0.3045
		9	0.1598	0.1292	0.1048
	Final Uniform Queue Length For Stream (pcu)	1	0.56	0.55	0.65
		2	1.49	1.58	1.86
		3	0.16	0.13	0.10
		4	3.06	2.82	2.87
		5	2.61	2.54	2.79
		6	0.70	0.56	0.45
		7	3.84	3.88	4.15
		8	3.31	3.26	3.39
		9	0.01	0.01	0.01
	Final Random Queue Length For Stream (pcu)	1	0.01	0.01	0.02
		2	0.48	0.58	0.65
		3	0.01	0.01	0.01
		4	1.07	0.87	0.62
		5	0.69	0.70	0.78
		6	0.40	0.35	0.31
		7	1.31	1.41	1.20
		8	1.92	1.83	1.27
		9	0.00	0.00	0.00
Total Delay (Pcu·Minutes) Estimated By	New Formula		**2635.23**	**2601.75**	**2844.05**
	OSCADY Formula		**2542.55**	**2511.08**	**2739.25**
Percentage Reduction In Total Delay Estimated By	New Formula			**1.27%**	**8.52%**
	OSCADY Formula			**1.24%**	**8.33%**

It can be seen from Table 5.19 that after the extension, the benefits arising from the sequential optimisation of signal settings are small (around 1.25%), but the results are still much better than that given by OSCADY. The general pattern of differences between the timings given by OSCADY and the new method is as discussed in Chapter 4.

§5.9.5 Discussions

From these few examples it can be seen that the new method of calculating signal settings for periods of time-varying demand can give somewhat better results than the program OSCADY. This can be explained by the fact that: (a) when the junction is oversaturated, OSCADY will maximise junction capacity rather than the total delay; (b) There is no interaction between the choices of signal settings in different periods in the optimisation process in OSCADY. The sequential optimisation method introduced in this Chapter, however, can take into account the interaction between successive time periods, by carrying forward the queue lengths left at the end of each time period, by sequential re-optimisation and by shifting the times at which the settings change. In these ways an improved performance index can be obtained. The results of the example calculations are consistent with the possibility that the new method may offer greater improvements in more complicated cases.

The example calculations in this Chapter have shown that it can be advantageous to optimise the traffic signal settings for periods of time-varying demand globally. A simple period-by-period optimisation will give only locally optimal solutions, which can be appreciably inferior to the solutions found by the more global heuristics reported here.

CHAPTER 6. CONCLUDING REMARKS AND SUGGESTIONS FOR FURTHER WORK

§6.1 CONCLUDING REMARKS

The literature review has shown that most of the existing methods of calculating settings for traffic signals at a single junction for a single time period are confined to equilibrium conditions. If the junction is oversaturated, optimisation with respect to delay will involve a non-convex, non-differentiable objective function. Even though some of the previous methods can deal with oversaturation, they find the signal settings either by optimising a different performance index instead of total delay, or by making some approximations when total delay is chosen as the objective function.

The development of the new delay expression introduced in §4.4 solves the difficulty of non-differentiability so that the derivatives can be obtained for use with gradient-based optimisation algorithms. The example calculations show that such a formula is appropriate for use, and even though such a delay formula is not strictly convex in the control variables, it is nearly convex and the non-convexity is unlikely to cause problems in finding the optimal signal settings.

Only the program OSCADY can model periods of time-varying demand, but it has some limitations: (a) when the junction is oversaturated, the program optimises the signal settings with respect to maximum junction capacity rather then total delay; (b) furthermore, the signal settings given by OSCADY are only optimal in each single time period but not for the whole succession of time-periods, hence are only local solutions.

In addition to using the new delay expression, the problem of optimisation over periods of time-varying demand has been reformulated to allow successive re-optimisation of the signal settings for the various periods with respect to total delay for the whole succession of time periods, together with optimisation of the times at which the settings of the signal change. The optimisation method, based on the subroutine OPTIM in the earlier procedure SIGSET, is capable of improving the junction control performance. The total delay will be reduced to a greater or better extent compared with the local solutions such as those given by OSCADY. In addition, the algorithm can take some account of delays subsequent to the time periods considered. Hence the problem of optimising traffic signals for periods of time-varying demand for a junction has partly been solved by the work described in this thesis. Further aspects that need to be considered are mentioned in the next section.

§6.2 SUGGESTIONS FOR FURTHER WORK

The work described so far in this thesis on the problem of setting traffic signals for periods of time-varying demand is only at its early stage and is incomplete, in a sense that there are a lot of potential areas to be further investigated so that more satisfactory results may be obtained. The most important ones are mentioned here.

§6.2.1 Further Investigations of Time-shifts

The algorithm for optimising the times at which the signal settings should change is based on the assumption that the optimal time for changing should be near to its initial value (often the time at which the level of demand is assumed to change, so that a search confined to the neighbourhood of the initial value will find the optimal time. Although this has been supported by all the example calculations for a two period cases, it is not completely clear whether it is possible for there to exist more than one minimum point even near the initial value, and more analysis and example calculations are necessary.

§6.2.2 Multi-criteria Optimisation

So far all the algorithms are based on minimising the total delay. Although this is a usual practice, there might exist some situation where optimisation with respect to other performance indices is necessary, or the performance index may be a weighted sum of total delay and fuel consumption, final queue lengths, total number of stops, or other relevant quantities. Such a multi-criteria optimisation method will offer wider choice for the users.

§6.2.3 Phase-based Optimisation

The methods developed in this thesis have been implemented in the stage-based approach. As mentioned in Chapter 1, this requires the pre-determination of the clearance times, the stage structure and the stage sequences, whereas in the phase-based approach, only the clearance times need to be specified on the basis of safety requirements; all other elements of the signal timings can be included in the optimisation procedure. Hence the phase-based method offers the potential opportunity for a better control performance by proper choice of the stage sequence and the interstage structures. The implementation of the present optimisation methodology in phase-based framework is therefore desirable. This presents no difficulties in principle, but because the number of control variables is bigger in the phase-based approach, it may take longer time for the present algorithms to run. An improved optimisation technique may be an alternative solution.

§6.2.4 Wider Experience

The work described here has shown that the differences between the signal settings and resulting delays and queue-lengths given by the new method and those given by existing methods can sometimes be very small, and in other cases substantial. Some basic reasons for such differences have been identified in the few examples studied here, but wider experience with the new method is needed to obtain a clearer understanding of the circumstances in which its use will offer appreciable advantages over existing methods.

REFERENCES

STAGE-BASED CONTROL METHODS

AKCELIK, R. (1982) New approximate expressions for delay, stop rate and queue length at isolated signals. Proc. Int. Conf. On Road Traffic Signalling, Inst. Elec. Engrs, London, 9-13.

AKCELIK, R. and BESLEY, M. (1984) SIDRA-2 user's guide. Australian Road Research Board Internal Report 367-5.

ALLSOP, R.E. (1970) Optimisation techniques for reducing delay to traffic in signalised road networks. Ph. D Thesis, University of London.

ALLSOP, R. E.(1971a) Delay-minimizing settings for fixed-time traffic signals at a single road junction. J. Inst. Maths Applics, 8(2), 164-185.

ALLSOP, R. E.(1971b) SIGSET: a computer program for calculating traffic signal settings. Traff. Engng & Cont.,13(2), 58-60.

ALLSOP, R. E.(1972a) Estimating the traffic capacity of a signalized road junction. Transpn Res., 6(3), 245-255.

ALLSOP, R.E. (1972b) Delay at a fixed time traffic signal-I: theoretical analysis. Transpn Sci., 6(3), 260-285.

ALLSOP, R. E. (1976) SIGCAP: a computer program for assessing the capacity of signal-controlled road junctions. Traff. Engng & Cont., 17(8/9), 338-341.

BRANSTON, D.(1978) A comparison of observed and estimated queue lengths at over-saturated traffic signals. Traff. Engng. & Cont., 19(7), 322-327.

BURROW, I.J. (1987) OSCADY: a computer program to model capacities, queues and delays at isolated traffic signal junctions. Transport and Road Research Laboratory Report RR105, Crowthorne.

CATLING, I. (1977) A time-dependent approach to junction delays. Traff. Engng. & Cont., 18(11), 520-523, 526.

CLAYTON, A.J.H. (1941) Road traffic calculations. J. Inst. Civ. Engrs 16, 247-284.

CRONJE, W.B. (1983) Optimization model for isolated signalized traffic intersections. Transportation Research Board TRR 905, 80-83.

CRONJE, W.B. (1986) Comparative analysis of models for estimating delay for oversaturated conditions at fixed-time traffic signals. Transportation Research Board TRR 1091, 48-59.

DE SMIT, J.H.A. (1971) Transient behaviour of the queues at a fixed cycle traffic light. Transpn. Res., 5, 1-14.

FEDERAL HIGHWAY ADMINISTRATION (1985) SOAP84 User's Manual. U.S. Department of Transportation, Washington DC.

GALLIVAN, S.(1982) A geometric proof that Webster's formula for a delay at a traffic signal is convex. Note from Transport Studies Group, University College London (unpublished).

HUTCHINSON, T.P. (1972) Delay at a fixed time traffic signal II: Numerical comparison of some theoretical expressions. Transpn Sci., 6(3), 286-305.

KIMBER, R.M., MARLOW, M. and HOLLIS,.E.M.(1977) Flow/delay relationship for major/minor priority junctions. Traff. Engng. & Cont., 18(11),516-519.

KIMBER, R.M. and HOLLIS, E.M.(1978) Peak period traffic delays at road junctions and other bottlenecks. Traff. Engng. & Cont., 19(10),442-446.

KIMBER, R.M and HOLLIS, E.M. (1979) Traffic queues and delays at road junctions. Transport and Road Research Laboratory Report LR909, Crowthorne.

MAYNE, A.J. (1976) Some improvements to programs for traffic signal settings control. Traff. Engng. & Cont., 17(6), 266-267.

MAYNE, A.J. (1979) "Traffic signal settings for congested traffic", In: Research Directions in Computer Control for Urban Traffic Systems. American Society of Civil Engineers, New York, 335 - 354.

MILLER, A.J. (1963) Settings for fixed-cycle traffic signals. Opl Res. Quarterly, 14, 373-386.

MILLER, A.J. (1968) The capacity of signalised intersections in Australia. Australian Road Research Board Bulletin No. 3.

MURCHLAND, J.D. (1977) The convexity of Webster's two-term traffic signal delay formula. Note from Transport Studies Group, University College London (unpublished).

NEWELL, G.F. (1965) Approximation methods for queues with applications to the fixed cycle traffic light. SIAM Rev., 7, 223-239.

OHNO, K. and MINE, H. (1973a) Optimal traffic signal settings-I. criterion for undersaturation of a signalized intersection and optimal signal setting. Transpn Res., 7(1), 243-267.

OHNO, K. and MINE, H. (1973b) Optimal traffic signal settings-II. a refinement of Webster's method. Transpn Res., 7(1), 269-292.

OHNO, K. (1978) Computational algorithm for a fixed cycle traffic signal and new approximate expressions for average delay. Transpn. Sci. 12(1), 29-47.

RELJIC, S. (1988) TRAFSIG: a computer program for signal settings at an isolated under- or oversaturated, fixed-time controlled intersection. Traff. Engng. & Cont. 29(11), 562-566.

WARDROP, J.G. (1952) Some theoretical aspects of road traffic research. Proc. Inst. Civ. Engng., 1(2), 325-378.

WEBSTER, F.V. (1958) Traffic signal settings. Road Research Technical Paper 39. HMSO, London.

WEBSTER, F.V. and COBBE, B.M. (1966) Traffic signals. Road Research Technical Paper 56. HMSO, London.

KIMBER, R.M., McDONALD, M and HOUNSELL, N (1986) The prediction of saturation flows for road junctions controlled by traffic signals. Transport and Road Research Laboratory Report RR67, Crowthorne.

GALLIVAN, S. and HEYDECKER, B. G.(1988) Optimising the control performance of traffic signals at a single junction. Transpn. Res., 22B(5), 357-370.

HEYDECKER, B.G. (1982) Generation of signal control sequences with double green intervals. Note from Transport Studies Group, University College London (unpublished).

HEYDECKER, B.G. and DUDGEON, I.W. (1987) Calculation of traffic signals to minimise delay at a junction. Proc. 10th Int. Symposium on Transpn. & Traff. Theory, Elsevier Science Publishing Co., Inc., N.Y.

IMPROTA, G. and CANTARELLA, G.E. (1984) Control system design for an individual signalized junction. Transpn Res., 18B(2),147-167.

MÖLLER, K. (1987) Calculation of optimum fixed-time signal programs. Proc. 10th Int. Symposium on Transpn. & Traff. Theory, Elsevier Science Publishing Co., Inc., N.Y.

MORTON, G.D. and SILCOCK, J.P. (1987) A phase-based calculation method for settings at a single-controlled junction. Proc. Engineering Foundation Conf. on Management and Control of Urban Traffic, Henniker, Mass, June 1987.

MORTON, G.D. (1987) SQGN: A computer program for generating stage sequences at a signal controlled junction - an investigation into its practical applications. Note from Transport Studies Group, University College London (unpublished).

SANG, A.P. and SILCOCK, J.P. (1989) SIGSIGN user manual. Steer Davies and Gleave Ltd and Transport Studies Group, University College London.

SIMMONITE, B.F. (1985) LINSIG: a computer program to aid traffic signal design and assessment. Traff. Engng. & Cont., 26(6), 310-315.

STOFFERS, K.E. (1968) Scheduling of traffic lights-a new approach. Transpn Res., 2(3), 199-234.

TULLY, I.M.Z. (1977) Synthesis of sequences for traffic signal controllers using techniques of the theory of graphs. Ph.D. Thesis, Report OUEL 1189/77. Oxford: University of Oxford.

TULLY, I.M.Z. and MURCHLAND, J.D. (1978) Calculation and use of the critical cycle time for a single traffic controller. Proc. PTRC Summer Annual Meet. PTRC-P152, 96-112.

MATHEMATICAL METHODS

LOVITT, W.V. (1939) Elementary theory of equations. Prentice-Hall, Inc., Englewood Cliffs, N.J.

LUENBERGER, D.G. (1984) Linear and nonlinear programming (2nd Ed.). Addison-Wesley Publishing Company, Inc.

BURDEN, R.L. and FAIRES, J.D. (1985) Numericl analysis, 3rd ed. Boston: Prindle, Weber & Schmidt.

APPENDIX 1: THE SOLUTION OF A CUBIC EQUATION

This appendix is based on the book *'Elementary Theory of Equations'* (**Lovitt** 1939), but some of the notations are different from the original text.

Consider the general cubic with real coefficients:

$$a_0X^3 + a_1X^2 + a_2X + a_3 = 0 \qquad (a_0 \neq 0) \qquad \text{(A1.1)}$$

Divided by a_0, (A1.1) can be written in the form:

$$X^3 + b_1X^2 + b_2X + b_3 = 0 \qquad \text{(A1.2)}$$

Where $b_i = a_i/a_0$, $i = 1, 2, 3$.

Let $X = Z - b_1/3$, then a *reduced cubic equation* can be obtained:

$$Z^3 + p_1Z + p_2 = 0 \qquad \text{(A1.3)}$$

We will only consider the case where $p_1 \neq 0$, $p_2 \neq 0$, which is the case for equation (4.12a). In other cases such as where at least p_1 or p_2 is zero, the solution is straightforward and will therefore not be discussed here.

Introduce two new variables A and B, and their respective cube roots, u,v, where

$$u^3 = A \text{ and } v^3 = B$$

Since a number has three cube roots, u and v each has three values. Let

$$Z = u + v \qquad \text{(A1.4)}$$

Then Z has nine values, but a cubic has only three roots, therefore we must place some restrictions on u and v so that Z will have three and only three values. Substitute (A1.4) in (A1.3), we have:

$$u^3 + v^3 + (3uv + p_1)(u+v) + p_2 = 0 \qquad \text{(A1.5)}$$

Equation (A1.5) can be simplified by imposing the condition on u and v that

$$3uv + p_1 = 0 \tag{A1.6}$$

Then $$u^3v^3 = -\frac{p_1^3}{27} \tag{A1.7}$$

And equation (A1.3) can be reduced to:

$$u^3 + v^3 = -p_2 \tag{A1.8}$$

From (A1.7) and (A1.8), u^3 and v^3 may be considered as the two roots, A and B, of the quadratic equation

$$Y^2 + p_2Y - \frac{p_1^3}{27} = 0 \tag{A1.9}$$

The solution for (A1.9) is:

$$Y = -p_2/2 \pm \sqrt{\Delta}$$

Where $\Delta = p_1^3/27 + p_2^2/4$

Hence we may set

$$\begin{aligned} u^3 &= -p_2/2 + \sqrt{\Delta} = A \\ v^3 &= -p_2/2 - \sqrt{\Delta} = B \end{aligned} \tag{A1.10}$$

$$\left.\begin{aligned} &\text{Then } u^3 = \sqrt[3]{A}\ ;\ \omega\sqrt[3]{A}\ ;\ \omega^2\sqrt[3]{A} \\ &\text{And } v^3 = \sqrt[3]{B}\ ;\ \omega\sqrt[3]{B}\ ;\ \omega^2\sqrt[3]{B} \end{aligned}\right\} \tag{A1.11}$$

where $\omega = -0.5 + 0.5i\sqrt{3}$ and $\omega^2 = -0.5 - 0.5i\sqrt{3}$

From (A1.6) the values of u and v in (A1.4) must be paired in such a way that $uv = -p_1/3$. There are only three pairs of values in (A1.11) satisfy this condition, these are:

$$\sqrt[3]{A}\,,\ \sqrt[3]{B}\ ;\ \omega\sqrt[3]{A}\,,\ \omega\sqrt[3]{B}\ ;\ \omega^2\sqrt[3]{A}\,,\ \omega^2\sqrt[3]{B}\,.$$

Hence from (A1.4) we get the three roots of (A1.3), namely

$$Z_1 = \sqrt[3]{A} + \sqrt[3]{B}\ ;\ Z_2 = \omega\sqrt[3]{A} + \omega^2\sqrt[3]{B}\ ;\ Z_3 = \omega^2\sqrt[3]{A} + \omega\sqrt[3]{B} \tag{A1.12a}$$

These roots can be discussed further according to the discriminate Δ.

1. If $\Delta = 0$ all roots are real. Two roots are equal

$$Z_1 = 3p_2/p_1,\ Z_2 = Z_3 = -3p_2/2p_1$$

2. If $\Delta > 0$ one root is real and two are imaginary.

From (A1.10), A and B are real and distinct. Hence from (A1.12a), Z_1 is real and Z_2 and Z_3 are conjugate imaginary numbers.

3. If $\Delta < 0$ all roots are real and distinct.

From (A1.10), A and B are conjugate imaginary numbers, Let $A = A_1 + A_2 i$ and $B = A_1 - A_2 i$. Then $\sqrt[3]{A}$ and $\sqrt[3]{B}$ will also be conjugate imaginary numbers, for suitable choice of the cube roots.

Let $\sqrt[3]{A} = B_1 + iB_2$ and $\sqrt[3]{B} = B_1 - iB_2$ $(B_2 \neq 0)$

The three roots will be:

$$Z_1 = 2B_1;\quad Z_2 = -B_1 - B_2\sqrt{3}\ ;\quad Z_3 = -B_1 + B_2\sqrt{3} \tag{A1.12b}$$

When $\Delta < 0$ the equation (A1.3) can also be solved by trigonometric method. Since we have the equality

$$\mathrm{Cos}3\theta = 4\mathrm{Cos}^3\theta - 3\mathrm{Cos}\theta$$

so that

$$\mathrm{Cos}^3\theta - 0.75\mathrm{Cos}\theta - 0.25\mathrm{Cos}3\theta = 0 \tag{A1.14}$$

In (A1.3), put $Z = Y/n$; then

$$Y^3 + p_1n^2Y + n^3p_2 = 0 \tag{A1.15}$$

and (A1.14) and (A1.15) will be identical if

$$Y = \mathrm{Cos}\theta;\ p_1n^2 = -0.75;\ n^3p_2 = -0.25\mathrm{Cos}3\theta.$$

i.e.

$$n = \sqrt{-3/(4p_1)}$$

and

$$\mathrm{Cos}3\theta = -4p_2(-0.75/p_1)^{3/2} = -0.5p_2(-27/p_1^3)^{1/2} \tag{A1.16}$$

These equations can always be solved if p_1 is negative, and

$$\left|\frac{p_2}{2}\left\{-\frac{27}{p_1^3}\right\}^{1/2}\right| < 1.$$

Since $\Delta < 0$ this condition is satisfied. If θ is the smallest value satisfying (A1.16), then the values $\theta + 2\pi/3$ and $\theta + 4\pi/3$ also satisfy it, so that the roots of the equation (A1.3) are

$$Z_1 = \frac{\mathrm{Cos}\theta}{n},\ Z_2 = \frac{\mathrm{Cos}(\theta+2\pi/3)}{n},\ Z_2 = \frac{\mathrm{Cos}(\theta+4\pi/3)}{n}$$

It can be proven that these three roots are the same as given by (A1.12b).

APPENDIX 2: THE DERIVATIVES OF THE NEW DELAY FORMULA

A2.1 The Expression For The New Delay Formula

Let the delay per unit time for a stream be denoted by D (for simplicity, no subscript is used here). Then according to the new delay formula (4.20), in terms of practical computing, D can be simplified and calculated by the following procedure:

$$R_0 = qT+2L_{r0}+Qc(1-\Lambda) \qquad \text{(A2.1)}$$

$$a_0 = \Lambda(QT-2C) \qquad \text{(A2.2)}$$

$$a_1 = 2C-Qc(1-\Lambda)^2-\Lambda R_0-QT(\Lambda+1) \qquad \text{(A2.3)}$$

$$a_2 = Qc(1-\Lambda)^2+R_0(\Lambda+1)+QT \qquad \text{(A2.4)}$$

$$a_3 = -R_0 \qquad \text{(A2.5)}$$

$$R_1 = a_1^2-3a_0a_2 \qquad \text{(A2.6)}$$

$$R_2 = 9a_0a_1a_2-2a_1^3-27a_0^2a_3 \qquad \text{(A2.7)}$$

1.If $a_0 \neq 0$:

$$B_1 = 0.5R_2/\sqrt{R_1^3} \qquad \text{(A2.8)}$$

$$\theta = \mathrm{ArcCos}(\pm B_1) = \begin{cases} \mathrm{ArcCos}B_1, & a_0>0 \\ \mathrm{ArcCos}(-B_1), & a_0<0 \end{cases} \qquad \text{(A2.9)}$$

$$B_2 = 2\sqrt{R_1}\ \mathrm{Sin}(\theta/3\pm\pi/6)+a_1 = \begin{cases} 2\sqrt{R_1}\ \mathrm{Sin}(\theta/3+\pi/6)+a_1, & a_0>0 \\ 2\sqrt{R_1}\ \mathrm{Sin}(\theta/3-\pi/6)+a_1, & a_0<0 \end{cases} \qquad \text{(A2.10)}$$

$$D = 0.5(R_0+QTB_2/3a_0) \qquad \text{(A2.11a)}$$

2. If $a_0=0$:

$$X_e = \frac{-a_2+\sqrt{a_2^2-4a_1a_3}}{2a_1} \qquad \text{(A2.11b)}$$

$$D = 0.5(R_0-QTX_e) \qquad \text{(A2.11c)}$$

A2.2 The Calculation Of The Derivatives

The delay per unit time defined by (A2.11) can be regarded as a function of Λ, λ_0, and L_{r0}, i.e. $D = D(\Lambda, \lambda_0, L_{r0})$. In this thesis, the derivatives of D with respect to these three variables are needed in the various optimisation routines.

1. $\partial D/\partial\Lambda$:

$$\frac{\partial(QT)}{\partial\Lambda} = sT \tag{A2.12}$$

$$\frac{\partial(R_0)}{\partial\Lambda} = sc(1-2\Lambda) \tag{A2.13}$$

$$\frac{\partial(a_0)}{\partial\Lambda} = 2(QT-C) \tag{A2.14}$$

$$\frac{\partial(a_1)}{\partial\Lambda} = -sc(1-4\Lambda+3\Lambda^2)-R_0-\Lambda\frac{\partial R_0}{\partial\Lambda} -sT(2\Lambda+1) \tag{A2.15}$$

$$\frac{\partial(a_2)}{\partial\Lambda} = sc(1-4\Lambda+3\Lambda^2)+R_0+(\Lambda+1)\frac{\partial R_0}{\partial\Lambda} +sT \tag{A2.16}$$

$$\frac{\partial(a_3)}{\partial\Lambda} = -\frac{\partial R_0}{\partial\Lambda} \tag{A2.17}$$

(1) If $a_0 \neq 0$:

$$\frac{\partial(R_1)}{\partial\Lambda} = 2a_1\frac{\partial(a_1)}{\partial\Lambda} - 3\left\{a_0\frac{\partial(a_2)}{\partial\Lambda} + a_2\frac{\partial(a_0)}{\partial\Lambda}\right\} \tag{A2.18}$$

$$\begin{aligned}\frac{\partial(R_2)}{\partial\Lambda} = 9\Bigg\{&\frac{\partial(a_0)}{\partial\Lambda}(a_1a_2-3a_0a_3)\\ &+ a_0\left(a_1\frac{\partial(a_2)}{\partial\Lambda}+a_2\frac{\partial(a_1)}{\partial\Lambda} - 3\left[a_0\frac{\partial(a_3)}{\partial\Lambda} +a_3\frac{\partial(a_0)}{\partial\Lambda}\right]\right)\Bigg\}\\ &- 6a_1^2\frac{\partial(a_1)}{\partial\Lambda}\end{aligned} \tag{A2.19}$$

$$\frac{\partial(B_1)}{\partial\Lambda} = \frac{1}{2R_1^3}\left\{\frac{\partial(R_2)}{\partial\Lambda}\cdot\sqrt{R_1^3} - \frac{3}{2}R_2\sqrt{R_1}\,\frac{\partial(R_1)}{\partial\Lambda}\right\} \tag{A2.20}$$

$$\frac{\partial\theta}{\partial\Lambda} = \mp \frac{\partial(B_1)}{\partial\Lambda}\cdot\frac{1}{\sqrt{1-B_1^2}} \tag{A2.21}$$

$$\frac{\partial(B_2)}{\partial\Lambda} = \frac{\partial(a_1)}{\partial\Lambda}+\frac{1}{\sqrt{R_1}}\cdot\frac{\partial(R_1)}{\partial\Lambda}\mathrm{Sin}(\theta/3\pm\pi/6)+\frac{2\sqrt{R_1}}{3}\cdot\frac{\partial\theta}{\partial\Lambda}\mathrm{Cos}(\theta/3\pm\pi/6) \tag{A2.22}$$

$$B_3 = QTa_0\frac{\partial(B_2)}{\partial\Lambda} + \left\{a_0\frac{\partial(QT)}{\partial\Lambda} - QT\frac{\partial(a_0)}{\partial\Lambda}\right\}B_2 \tag{A2.23}$$

$$\frac{\partial D}{\partial\Lambda}\ 0.5\left\{\frac{\partial(R_0)}{\partial\Lambda} + \frac{B_3}{3a_0^2}\right\} \tag{A2.24a}$$

(2) If $a_0=0$:

$$\frac{\partial X_c}{\partial\Lambda} = \frac{0.5\left\{-\frac{\partial(a_2)}{\partial\Lambda} + \frac{1}{2\sqrt{a_2^2-4a_1a_3}}\left(2a_2\frac{\partial(a_2)}{\partial\Lambda} - 4\left[a_1\frac{\partial(a_1)}{\partial\Lambda}+a_3\frac{\partial(a_3)}{\partial\Lambda}\right]\right)\right\}}{a_1}$$

$$-\frac{X_c}{a_1}\frac{\partial(a_1)}{\partial\Lambda} \tag{A2.24b}$$

$$\frac{\partial D}{\partial\Lambda} = 0.5\left\{\frac{\partial(R_0)}{\partial\Lambda} - \left\{X_c\frac{\partial(QT)}{\partial\Lambda} + QT\frac{\partial(X_c)}{\partial\Lambda}\right\}\right\} \tag{A2.24c}$$

2. $\partial D/\partial\lambda_0$:

$$\frac{\partial(QT)}{\partial\lambda_0} = 0 \tag{A2.25}$$

$$\frac{\partial(R_0)}{\partial\lambda_0} = -Qc(1-\Lambda)/\lambda_0 \tag{A2.26}$$

$$\frac{\partial(a_0)}{\partial\lambda_0} = 0 \tag{A2.27}$$

$$\frac{\partial(a_1)}{\partial\lambda_0} = \frac{Qc(1-\Lambda)^2}{\lambda_0} - \Lambda\frac{\partial(R_0)}{\partial\lambda_0} \tag{A2.28}$$

$$\frac{\partial(a_2)}{\partial\lambda_0} = -\frac{Qc(1-\Lambda)^2}{\lambda_0} + (\Lambda+1)\frac{\partial(R_0)}{\partial\lambda_0} \tag{A2.29}$$

$$\frac{\partial(a_3)}{\partial\lambda_0} = -\frac{\partial(R_0)}{\partial\lambda_0} \tag{A2.30}$$

Then follow the steps (A2.18)-(A-2.24) except that $\partial\Lambda$ is replaced by $\partial\lambda_0$.

3. $\partial D/\partial L_{r0}$:

$$\frac{\partial(QT)}{\partial L_{r0}} = 0 \tag{A2.31}$$

$$\frac{\partial(R_0)}{\partial L_{r0}} = 2 \tag{A2.32}$$

$$\frac{\partial(a_0)}{\partial L_{r0}} = 0 \tag{A2.33}$$

$$\frac{\partial(a_1)}{\partial L_{r0}} = -\Lambda\frac{\partial(R_0)}{\partial L_{r0}} = -2\Lambda \tag{A2.34}$$

$$\frac{\partial(a_2)}{\partial L_{r0}} = (\Lambda+1)\frac{\partial(R_0)}{\partial L_{r0}} = 2(\Lambda+1) \tag{A2.35}$$

$$\frac{\partial(a_3)}{\partial L_{r0}} = -\frac{\partial(R_0)}{\partial L_{r0}} = -2 \tag{A2.36}$$

Then follow the steps (A2.18)-(A-2.24) except that $\partial\Lambda$ is replaced by ∂L_{r0}.

APPENDIX 3: THE DERIVATIVES OF THE RANDOM QUEUE LENGTH FORMULA

A3.1 The Expression For The Random Queue Length Formula

As pointed out in Chapter 3, when the derivatives of the random queue length formula with respect to the control variables need to be evaluated, the formula such as (3.7) that can be used to estimate the random queue length with more accuracy cannot be used for this purpose, and formula (3.5b) should be employed. Numerically, for a traffic stream, the random queue length L_r at time T can be calculated by the following procedure:

$$A_1 = 2C(qT+L_{r0})^2 \tag{A3.1}$$

$$A_2 = (qT+L_{r0}-\Lambda sT)^2 + 4C(qT+L_{r0}) \tag{A3.2}$$

$$A_3 = (\Lambda sT)^2 + (2C-\Lambda sT)(qT+L_{r0}) \tag{A3.3}$$

Then.

$$L_r = \frac{A_1}{\Lambda sT\sqrt{A_2} + A_3} \tag{A3.4}$$

A3.2 The Calculation Of The Derivatives

The random queue length defined by (A3.4) can be regarded as a function of Λ and L_{r0}, i.e. $L_r = L_r(\Lambda, L_{r0})$. In this thesis, the derivatives of L_r with respect to these two variables are needed in the various optimisation routines.

1. $\partial L_r/\partial \Lambda$:

$$\frac{\partial(A_1)}{\partial \Lambda} = 0 \tag{A3.5}$$

$$\frac{\partial(A_2)}{\partial \Lambda} = -2(qT+L_{r0}-\Lambda sT)sT \tag{A3.6}$$

$$\frac{\partial(A_3)}{\partial \Lambda} = (2\Lambda sT-qT-L_{r0})sT \tag{A3.7}$$

Then

$$\frac{\partial L_r}{\partial \Lambda} = \frac{\frac{\partial(A_1)}{\partial \Lambda}\cdot\left[\Lambda s T\sqrt{A_2} + A_3\right] - A_1\left[sT\sqrt{A_2} + 0.5\Lambda s T\frac{\frac{\partial(A_2)}{\partial \Lambda}}{\sqrt{A_2}} + \frac{\partial(A_3)}{\partial \Lambda}\right]}{\left[\Lambda s T\sqrt{A_2} + A_3\right]^2}$$

(A3.8)

2. $\partial L_r/\partial L_{r0}$:

$$\frac{\partial(A_1)}{\partial L_{r0}} = 4C(qT+L_{r0})-2\Lambda \qquad \text{(A3.9)}$$

$$\frac{\partial(A_2)}{\partial L_{r0}} = 2(qT+L_{r0}-\Lambda sT) + 4C \qquad \text{(A3.10)}$$

$$\frac{\partial(A_3)}{\partial L_{r0}} = 2C-\Lambda sT \qquad \text{(A3.11)}$$

Then

$$\frac{\partial L_r}{\partial L_{r0}} = \frac{\frac{\partial(A_1)}{\partial L_{r0}}\cdot\left[\Lambda s T\sqrt{A_2} + A_3\right] - A_1\left[0.5\Lambda s T\frac{\frac{\partial(A_2)}{\partial L_{r0}}}{\sqrt{A_2}} + \frac{\partial(A_3)}{\partial L_{r0}}\right]}{\left[\Lambda s T\sqrt{A_2} + A_3\right]^2}$$

(A3.12)