UNIVERSITY OF LEEDS
Institute for Transport Studies

ITS Working Paper 349 *ISSN 0142-8942*

January 1992

A REVIEW OF TRAFFIC SIGNAL CONTROL

SP Shepherd

This work was undertaken on a project sponsored by
the Science and Engineering Research Council (Grant Ref: GR/E/38184)

Project title: Gating for traffic signal control: the application of state-space control theory.

CONTENTS

		Page
1.	INTRODUCTION	1
	1.1 Objectives of the paper	1
	1.2 Introduction to state space control theory	1
	1.2.1 Definitions	1
	1.3 Applications of state space in transportation	3
	1.4 Introduction to traffic control	3
	1.4.1 Objectives for congestion control	5
	1.4.2 Approaches to congestion control	5
	1.4.3 Types of congestion	6
2.	STATIC/PRE-PLANNED MEASURES FOR CONGESTION CONTROL	8
	2.1 Congestion pricing/road pricing	8
	2.2 Pre-planned signal settings	9
	2.3 Coordination of traffic signals	9
	2.4 Traffic metering and traffic restraint	11
3.	THEORETICAL STUDIES	12
4.	EXISTING TRAFFIC CONTROL SYSTEMS	21
	4.1 Dynamic control measures VERON and STAUKO	22
	4.1.1 Control program VERON	22
	4.1.2 STAUKO	23
	4.2 An expert system approach to congestion: SAGE	24
	4.3 SCOOT	24
	4.4 OPAC	26
	4.5 UTOPIA	27
	4.5.1 Basic UTOPIA concepts	27
	4.5.2 The dynamic traffic model	29
	4.5.3 The control	31
	4.6 PRODYN	32
	4.6.1 Control strategy	32
	4.7 SCATS	34
5.	CONCLUSIONS AND RECOMMENDATIONS	34
6.	REFERENCES	36

A REVIEW OF TRAFFIC SIGNAL CONTROL

1. INTRODUCTION

1.1 OBJECTIVES OF THE PAPER

The aim of this paper is to provide a starting point for the future research within the SERC sponsored project "Gating and Traffic Control: The Application of State Space Control Theory".

It will provide an introduction to State Space Control Theory, State Space applications in transportation in general, an in-depth review of congestion control (specifically traffic signal control in congested situations), a review of theoretical works, a review of existing systems and will conclude with recommendations for the research to be undertaken within this project.

1.2 INTRODUCTION TO STATE SPACE CONTROL THEORY

The state space method is based on the description of system equations in terms of n first-order difference equations or differential equations, which may be combined into a first-order vector-matrix difference equation or differential equation. The use of the vector-matrix notation greatly simplifies the mathematical representation of the systems of equations.

System design by use of the state space concept enables the engineer to design control systems with respect to given performance indices. In addition, design in the state space can be carried out for a class of inputs, instead of a specific input function such as the impulse function, step function, or sinusoidal function. Also, state space methods enable the engineer to include initial conditions in the design; which is not possible with conventional design methods.

1.2.1 Definitions

STATE
The state of a dynamic system is the smallest set of variables (called state variables) such that the knowledge of these variables at $t = t_{0\_}$ together with the knowledge of the input for $t>t_{0\_}$ completely determines the behaviour of the system for any time $t>t_{0\_}$.

STATE VARIABLES
The state variables of a dynamic system are the variables making up the smallest set of variables which determine the state of the dynamic system.

Note that state variables need not be physically measurable or observable quantities. Variables which do not represent physical quantities and those which are neither measurable nor observable can be chosen as state variables. Such freedom in choosing state variables is an advantage of the state space methods. Practically speaking, however, it is convenient to choose easily measurable quantities for the state variables, if this is possible at all, because optimal control laws will require the feedback of all state variables with suitable weighting.

STATE VECTOR

If n state variables are needed to completely describe the behaviour of a given system, then these n state variables can be considered the n components of a vector X. Such a vector is called a state-vector.

STATE SPACE

The n dimensional space whose coordinate axes consist of the - x_1 axis, x_2 axis, ..., x_n axis is called a state space. Any state can be represented by a point in the state space.

State Space Equations

For time varying (linear or non-linear) discrete time systems, the state equation may be written as

$$x(k+1) = f\,[\,x(k), u(k$$

and the output equation as

$$y(k) = g\,[\,x(k), u(k$$

For linear time-varying discrete-time systems, the state equation and output equation may be simplified to

$$x(k+1) = G(k)x(k) + H(k)u$$

and

$$y(k) = C(k)x(k) + D(k)u$$

where

x(k)	=	n-vector	(state vector)
y(k)	=	m-vector	(output vector)
u(k)	=	r-vector	(input vector)
G(k)	=	n x n matrix	(state matrix)
H(k)	=	n x r matrix	(input matrix)
C(k)	=	m x n matrix	(output matrix)
D(k)	=	m x r matrix	(direct transmission matrix).

The appearance of the variable k in the arguments of matrices G(k), H(k), C(k) and D(k) implies that these matrices are time varying. If the variable k does not appear explicitly in the matrices, they are assumed to be time-invariant, or constant. That is, if the system is time-invariant, then the last two equations can be simplified to

$$\cdot 1) = G\,x(k) + Hu(k) \qquad \textbf{(1)}$$

$$) = Cx(k) + Du(k) \quad \textbf{(2)}$$

similarly linear time-invariant continuous-time systems may be represented by the following state equation and output equation:

$$)= Ax(t)+ Bu(t) \quad \textbf{(3)}$$

$$)= Cx(t)+ Du(t) \quad \textbf{(4)}$$

Figure 1(a) shows the block diagram representation of the discrete-time control system defined by equations (1) and (2), and Figure 1(b) shows the continuous-time control system defined by equations (3) and (4). Notice that the basic configurations of the discrete-time and continuous-time systems are the same.

1.3 APPLICATIONS OF STATE SPACE IN TRANSPORTATION

The use of state space has been evident in various fields of transportation. It has been used to describe the dynamic equilibrium traffic assignment problem (Wie, 1989), to control the lateral movements of vehicles (Tomezuka and Peng, 1989); (Wei-bin Zhang, Parsons, 1989); (Sheikholeslam and Desoer, 1989); (P K Sinha, 1977).

State space has also been used as the basis for many theoretical studies such as those carried out by Papageorgiou et al (1990), Cremer and Schoof (1989), D'Ans and Gazis (1976), where it has been used as a model to describe various traffic characteristics and problems. These and other theoretical studies will be discussed later.

State space representation is also the basis of some existing traffic control systems such as UTOPIA (Mauro et al, 1984), OPAC (Gartner, 1989), and PRODYN (Henry and Farges, 1989). These and other existing traffic control systems will also be discussed in detail later.

1.4 INTRODUCTION TO TRAFFIC CONTROL

Urban vehicular traffic, as an expression of human behaviour, is variable in time and in space. The control of such traffic requires a high degree of adaptiveness to enable a suitable response to this variability. Ever since the inception of modern traffic signal controls, traffic engineers and signal system designers have attempted to make them as responsive as possible to prevailing traffic conditions. The premise always was that increased responsiveness must lead to improved traffic performance. This premise was applied to single intersection signals as well as to arterial and network signal systems. However, the extent to which traffic responsiveness is achieved depends on a variety of factors, which include: strategy sophistication, hardware capabilities, surveillance and communication equipment, as well as the operator capabilities.

With the advent of computerised systems in the mid 1960's, many cities embarked on deploying centrally controlled and monitored traffic signal systems. Such systems offered significant advantages when compared with the previously used electromechanical devices. However, they also imposed rigidities which restricted the opportunities for traffic responsiveness. Various aspects of these rigidities still persist today.

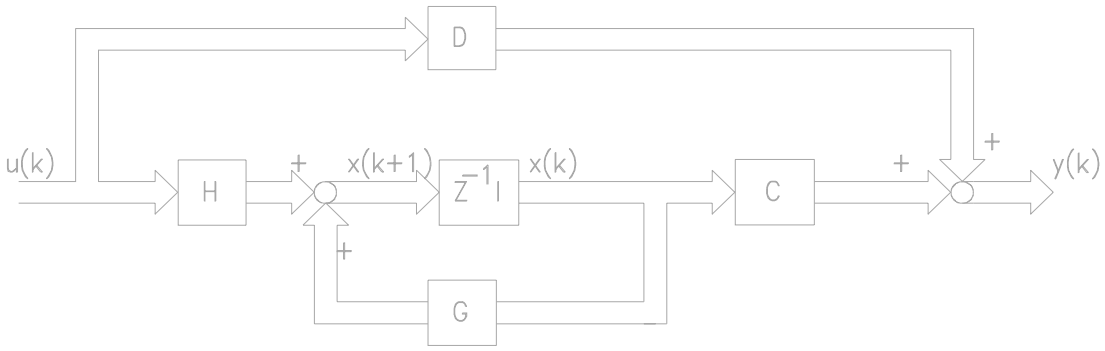

Figure 1 (a) Block diagram of the linear-time invariant discrete-time control system represented in state space

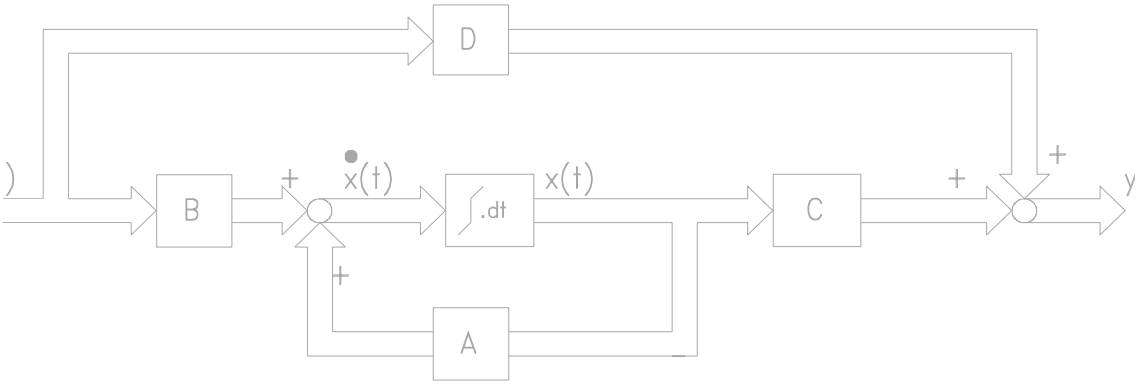

Figure 1 (b) Block diagram of the linear-time invariant continuous-time system represented in state space

The advent of microprocessor technologies drastically changed the traffic signal control field and opened up new horizons and opportunities for demand-responsive control. It became feasible to develop much more sophisticated systems than before, systems that offer a great deal of responsiveness, work largely automatically and eliminate the need for operator intervention.

1.4.1 Objectives for congestion control

In the past control measures aimed to minimise delay for all vehicles using the road system. With increasing concern in the green/environmental issues a range of objective functions must be put forward:

- minimise overall delay to vehicles
- minimise delays to public transport
- minimise delays to emergency services
- minimise delays to pedestrians
- equitable distribution of delays between competing traffic
- maximise reliability, ie. minimise unpredictable variations in journey time for vehicle users
- maximise network capacity
- minimise accident potential for all users
- minimise environmental impact of vehicular traffic (noise, atmospheric pollution, visual intrusion)
- energy efficiency.

It is important to note that some of the objectives do conflict and a compromise may have to be made in the selection of objectives.

However, some objectives can be met in tandem, for example minimising delay to vehicles would also help to minimise fuel consumption, atmospheric pollution and increase network throughput.

1.4.2 Approaches to congestion control

At the present most signal control methods aim to ease or prevent congestion. There are few papers which deal directly with queue management in overloaded conditions. Schemes can be put into three main groups:

i)Capacity improvements: critical queues can be reduced by eliminating or relocating bottlenecks.
ii)Queue regulation: traffic input to the tail of the critical queue can be regulated in order to control the queue length.
iii)Restricting traffic input area-wide: traffic entering a busy area can be metered or restrained so as to reduce the probability of critical queue formation anywhere in the area.

Junction improvements and road widening schemes can eliminate bottlenecks and hence reduce critical queue lengths. However, these measures are expensive and not always feasible. Many bottlenecks are associated not so much with the physical lay-out of the road network, but with pedestrian and commercial activities which spill into the carriageway. Measures to control parking, loading and pedestrian crossing movements are frequently applied, but they are often inconvenient to other users.

Improvements have been made through efficient traffic signal control methods. These have

advantages over other capacity measures in that they are flexible, cheap and cause less disturbance to the urban fabric.

Re-configuring the traffic routeing pattern within a city can increase capacity. For example, a one-way system can re-distribute conflicts and exploit unused stop-line capacity on adjacent streets.

Any capacity expansion strategy will promote further traffic growth. Consequently, it may postpone the problem - and even make it worse - rather than providing a long-term solution.

Queue control measures, on the other hand, involve regulating input, as opposed to boosting output (although it may have the effect of increasing capacity of the network, through minimising interference). By regulating the arrival rate and hence reducing the length of a critical queue, one can effectively move it to a new site, chosen to minimise the likelihood of queue propagation, or disruption as it propagates.

Measures in the third main group are based on the same general principle as the second; the distinction is that they are applied on an area-wide scale. The main justification for restraint is that it releases road space for other purposes, or contributes in some other way to wider objectives such as safety, urban conservation, or environmental quality.

Further to these three groups, measures can be classified as either "static" (pre-planned) or "dynamic" in operational terms.

1.4.3 Types of congestion

Traffic congestion occurs on all road environments in one form or another. First there is the slow crawl along busy shopping streets, then the long queues that form when accidents occur on rural roads, and the rapid, unpredictable slowing down and speeding up of traffic on overloaded motorways (often without any obvious physical cause).

However, the most acute form of traffic congestion is the area-wide "jam", in which vehicles can become embedded for long periods of time.

It is usually triggered by the interaction between two or more queues.

Traffic Queues

Jams start with queues. Huddart and Wright (1989) define the queueing mechanism as the basis for understanding "jams".

Queueing occurs when the capacity of the road system falls short of the demands placed on it by the traffic arriving at that time. The cause may be due to the following:

- external sources of "friction" such as pedestrians and parked vehicles;
- internal sources of "friction" such as vehicles turning into and out of side streets and access ways, and slow vehicles such as heavy lorries, buses and cyclists;
- isolated pinchpoints where the physical lay-out of the road effectively reduces capacity to a value less than that for neighbouring locations: typically these occur at junctions, and at narrow sections of road;
- instability in the traffic stream as flow approaches capacity.

All these can be referred to as "bottlenecks". On any road containing a bottleneck, traffic speeds tend to fall as the flow increases. The effect can be analysed as a queueing mechanism: passage through the bottleneck is pictured as a "service" which requires a given amount of time for each vehicle. If the upstream flow varies randomly from moment to moment, there will be periods when more vehicles arrive than can be serviced, and these vehicles must form a queue. During other periods the demand will be less than the service rate, and any queue present will diminish.

Over a long period of time, the average queue length will be determined according to the ratio of flow to capacity. As this ratio approaches unity, the number of vehicles in the queue (and hence the average delay per vehicle) will increase without limit.

Bottlenecks may be difficult to identify, they may be temporary obstructions which occur at different times and places on different days. They include crossing pedestrians, parked and loading vehicles, turning traffic and driver behaviour.

In these circumstances the traffic cannot be modelled as a simple queueing mechanism with only a single server. It is usual to express traffic behaviour in terms of empirically determined relationships between average speed, flows and concentration.

Note that queueing theory assumes that the service characteristics are not affected by changes in demand. At high levels of demand movement through a bottleneck may breakdown. The flow may actually fall below the level previously sustained through the bottleneck before breakdown. Huddart and Wright (1989) present two cases where this effect can be observed. First, on a high-speed road, if vehicles travel close together, any reduction in speed by one driver will cause successive drivers to brake more sharply. As the intensity of the braking increases, from vehicle to vehicle so the state either side of the shock wave will change until the traffic stream eventually comes to a halt.

Individual vehicles will encounter this "Stopping Shockwave", slow down to a crawl, and eventually clear the front of the queue via a "starting wave", which also propagates from vehicle to vehicle within the traffic stream. Between the stopping wave and the starting wave, there will be an almost stationary queue of traffic. Due to these different states and lags as vehicles accelerate, the flow discharging from the front of the queue will be less than the flow arriving at the rear. Consequently, the capacity of the lane is effectively reduced.

The second case occurs on narrow two-way streets in urban areas. If one direction of flow is obstructed by a parked vehicle, the unobstructed stream must occasionally give way so that the opposing traffic can pass.

Suppose a second vehicle parks nearby. There will now be two bottlenecks, each of single lane width, with a pocket of clear two-way road between. Now, if vehicles from the two traffic streams enter these two bottlenecks simultaneously, the system will "lock-up" so that no-one can move.

In the first case, the traffic has entered a "congested flow regime" where the speed and flow fall to zero. The transition from the free-flow regime to the congested-flow regime is unstable: it cannot be reversed without some form of external intervention. In the case of the saturated motorway lane, this would mean reducing the supply of vehicles, whereas in the case of the urban street, one or more vehicles would have to reverse out of one of the bottlenecks. Both speed and flow are limited by the physical characteristics of the bottlenecks.

Traffic Characteristics and Measured Variables

First it must be recognised that traffic represents the wishes and habits of the human beings and the vehicles controlled by them. It is therefore a stochastic process, ie. it is composed of a very large number of single events caused by individuals and which occur more or less randomly. The result is that variables measured from traffic generally exhibit large scale random fluctuations which have to be smoothed to a certain extent in control processes.

The degree of smoothing and the measurement interval influence the response of the control system. Some smoothing processes have been developed which are more sensitive to rising traffic measurements, thus the control process reacts more quickly to the build up during peak loads.

A vast number of measurements are now available or will become available in the near future. Measured variables include vehicle counts (pulses), the occupancy periods of detectors (pulse duration), the gaps between consecutive vehicles and, when double detectors are installed, the travelling time for one specific route section (1/speed) and the vehicle length/type. Latest developments of micro processor detectors makes it possible to measure speed and type of vehicle from one detector loop only. The variables can be measured in terms of time intervals or for single vehicles. By further processing these measured values, measurements of the spatial density of vehicles, of waiting and delay times, of the number of stops, of queue lengths (traffic jams), of the distribution of the mean traffic intensity within a signal cycle and of the degree of saturation can also be obtained.

2. STATIC/PRE-PLANNED MEASURES FOR CONGESTION CONTROL

These measures are not responsive to traffic, they may be signal plans based on average flows or physical changes to the road lay-out or road markings.

2.1 CONGESTION PRICING/ROAD PRICING

Both these methods recognise that congestion is now acting as a traffic regulator. This is wasteful, costly and damaging to economic growth, the environment and safety.

Paying for the use of roads in congested areas would hopefully free the roads for those who pay and provide a better public transport system, through investment for those who change mode (dependant on policy).

The method of payment can take several different forms. Area licensing has been introduced in Singapore, where the drivers of low-occupancy vehicles have to buy special licences in order to enter the control area. In Holland drivers must pay tolls at points on the way into the city. Another possibility is to impose charges at various points within the network, using electronic detectors coupled to a central computer. This was shown to be technically feasible in Hong Kong, but was rejected because it invaded the users' privacy. Current disadvantages with road pricing are:

i) lack of privacy to users (in the case of Hong Kong);
ii) single price on entry - not dependant on mileage;
iii) expensive to implement.

Oldridge and Hunter (1989) put forward a new concept for pricing congestion. It is based upon the principle that when a driver is caught in congestion (anywhere), an in-car electronic device will detect the speed driven and the number of stops made. It will then charge the driver accordingly by decrementing a pre-paid smart-card. The advantages of this system would be that it applies anywhere in the country, it is cheap to implement; no street furniture or changes to the infrastructure are required. Also free-flow travel is free, it does not invade privacy and it claims to make the car thief-proof!

One of the main disadvantages is that it may change driver behaviour ie. encourage speeding or dangerous driving in urban areas.

2.2 PRE-PLANNED SIGNAL SETTINGS

At first signal plans were developed on the basis of average flows so as to minimise delay over all the approaches. Usually more than one plan was developed for each intersection so that plans could be switched in to try and cope with higher peak demands. As demand increases these delay based plans increase the cycle time which produces larger queues and spillbacks.

The objective of minimising delay itself may not be appropriate. It has been shown by Smith (1988) that if one or more of the junctions in the network become saturated, it is better to maximise capacity rather than attempting to minimise delay. Usually, this means providing extra green time for the approaches having the greatest saturation flows.
Longer queues then have to be stored on the disadvantaged approaches, which may not always be possible.

The way in which a capacity-maximising plan is incorporated into the daily plan sequence is important. Gazis and Potts (1965) worked out a schedule for an isolated junction. A minimum delay algorithm should be used during off-peak, changing to a maximum capacity algorithm as soon as the signal becomes saturated, ie. the queues on one or more approaches consistently fail to clear. At the peak of demand, a third algorithm should be introduced to clear queues from the disadvantaged approaches, the aim being to allocate capacity so that all queues vanish simultaneously. Finally, the minimum delay algorithm should be reintroduced.

2.3 COORDINATION OF TRAFFIC SIGNALS

Under relatively light traffic conditions in linear corridor systems, it is usual to try and coordinate sets of signals to produce a "Green Wave". This means that traffic passing through one set of signals meets subsequent signals at green. Hence the green period at any one junction is timed to start later than its predecessor by an amount roughly equal to the travel times of vehicles between the junctions. This is a progressive "offset" determined by free-flow conditions.

However in congested conditions, or due to turning traffic, there is inevitably an initial queue at the start of green and the "wave" is disrupted. This leads to the idea of "reverse" signal progression. The offset is set in such a way that the downstream green period starts before the upstream, thus giving the queue a chance to clear. The general aim being that the first vehicle to leave the upstream junction should meet the last vehicle in the downstream queue as it is setting off. These theories will be discussed later in the on-line optimisation of offset techniques.

However this is not the only way offsets may be utilised. May et al (1988) experimented with offsets in Bangkok, where outdated signal settings could not cope with the peak hour demand.

This resulted in the use of Police Control at major junctions. The approaches were running at near 100% saturation, queue prediction was limited and dependent on fluctuations in demand and incidents. When the old signals were in use queues blocked back to the upstream junction causing disruption to the cross-street movements and inducing illegal red-running. The main objectives of the queue management strategy were as follows:

- each vehicle expects a delay of one cycle at each junction
- queues should be as short as possible subject to downstream, vehicles should join the queue as it moves off
- if blocking back occurs it should be timed so as not to intrude upon cross-street movements
- starting and stopping waves should be controlled so that the stopping wave arrives after the green ahead and the starting wave a few seconds before the end of green ahead. This would discourage illegal red-running
- the stage lengths should be adjusted to take blocking into account ie. increase the effective red time.

The cycle time at certain junctions had to be doubled because the waves interacted not only at the upstream junction but also at the second upstream junction (ie. the starting and stopping waves converged 2 junctions upstream).

With all these objectives in mind, a modified version of TRANSYT 8 was run and the results were implemented. The network tested was made up of two one-way arterials and five cross-streets. The results were as desired, traffic was halted during the upstream green period, however it reduced red-running and reduced disruption to crossing movements. This method tackled current congestion through controlling the offsets on the main arterials. Tests on a ten junction network in Bangkok showed that the fixed time controls could achieve up to 10% greater system speeds than obtained by manual police control.

A similar approach, this time controlling cross-streets was made by Rathi (1988) in New York. The project was carried out in the 5th Avenue area of New York, basically one-way streets in a grid network. The main arterials ran North-South and the cross-streets East-West. The existing situation was a two-phase 90 second cycle with offsets set for progressive movements along the N-S arterials which had 55-60% of the green time. The short cross-street links, combined with high demand, parking and pedestrians to cause recurring spillback. This spillback disrupted the arterial offsets. It was noted that the offsets did not work with queues of 100 feet or more. Vehicles had to stop which created a new shock wave, affecting driver behaviour and driver perceptions in such a way as to reduce the effective capacity.

The objectives were to avoid spillback occurrence or at least decrease its frequency by setting the offsets and green splits on the cross-streets. Objectives for the arterial N-S was to eliminate effective red caused by spillback and to fully utilise every second of green time.

A mathematical model was developed which predicted the probability of spillback on the cross-streets. Using the model to reduce the probability of spillback resulted in the flaring of greens along the cross-streets. This "flared green" progression has the effect of "metering" the number of vehicles released from the upstream signal so that the downstream junction clears. This protects the storage area between junctions, ensures both green periods are fully used and reduces spillback occurrence. The offsets running N-S were set to zero, leaving the cross-street offsets free to be set. This zero offset and the alternate one-way cross-streets produce an unusual North-South arterial green progression.

When implemented the scheme was successful in that it reduced spillback and increased throughput of the system both N-S and E-W. This method requires further research if it is to be applied to two-way streets without a critical direction.

2.4 TRAFFIC METERING AND TRAFFIC RESTRAINT

It may be possible to improve conditions at a bottleneck by "gating" or "metering" traffic input at a suitable point upstream, so that the demand does not reach the critical level at which flow breakdown takes place. The flared green approach above is an example of this on a local level.

Rathi (1991) also simulated the effect of external metering, through traffic signal control, on New York city's central business district (CBD). The TRAF-NETSIM simulation model was used to evaluate the impact of external metering on traffic operations in the control area. Traffic performance under existing conditions (ie. without any metering control) was compared with performance under conditions with different rates (and location) of metering implemented for traffic entering the control area. Analyses were performed for morning peak-period traffic conditions.

Three different strategies were analysed:

i)Uniform metering at all entry points of the control area. The same level of metering is implemented throughout the periphery of the control area.
ii)Variable metering for the control area, ie. different levels of metering at different metering locations.
iii)Metering traffic entering from the Queensboro Bridge entrances only.

Restrictive and permissive metering rates were implemented under each of these three strategies. Metering was implemented directly by modifying input traffic volumes at all entry points of the control area to the desired in flow rate. Under metering control strategies, regardless of how they might be implemented, only the maximum possible in flow rate can be specified as a policy; the actual in flow rate depends on traffic conditions within the control area. Specifically, congested conditions within the control area can produce queues that limit the rate of traffic in flow below that permitted by the metering policy.

The study showed that external metering can improve traffic conditions in the control area, over various measures of effectiveness. However, the strategy chosen depends upon the current state of the control area compared to the desired state. A further detailed study is planned to identify optimal metering policies and procedures for real-life implementation.

In principle the metering technique could be applied over a large area of the network, smoothing out momentary peaks in demand. However, while it may protect a centre from the influx of morning peak traffic, it is hard to see how it can be implemented during the evening peak, when much of the traffic originates within the controlled area.

Traffic restraint aims to restrict throughput to a level significantly less than the current demand and capacity. In order for it to work, the extra delays imposed must be large enough to affect motorists' behaviour, either in terms of route choice, time of travel, mode of travel, or choice of origin or destination.

It is recognised that restraining traffic by imposing extra delays is necessarily inefficient. The network runs at low average speeds and journey times become unpredictable. Commuters are prepared to accept delays rather than transfer to public transport.

3. THEORETICAL STUDIES

There have been many theoretical studies over the years, ranging from studies of a single intersection through to studies of whole networks including freeway, ramp metering, route guidance and signal control. The papers discussed in this section give a broad view of traffic modelling and can be classified into three main groups:-

i) Traditional Min/Maximisation problems.
ii)Strategies for Over-Saturation.
iii)Total Network Models.

One of the most referenced papers in this area is that by D'Ans and Gazis (1976) "Optimal Control of Oversaturated Store-and-Forward Transportation Networks". A general method for the dynamic optimisation of multi-commodity flows in an over-saturated, store-and-forward transportation network was developed. The network is characterised by time-varying inputs, and the existence of queues in front of certain nodes where the incoming flows exceed the capacity to discharge them, during a certain period of oversaturation.

The network is modelled in State-space. The model consists of a set of state-space differential equations which describe firstly a two-node, seven arc network with no turns and is later extended to cover more general networks. The state variables are the vehicle queues and the time dependent flow rates on the approaches. The vehicle flows are controlled by a traffic light at each intersection. The cycle length is fixed and identical for each intersection, so that the only control variables are the two parameters u and v which are proportional to the main green phases at each intersection.

Two control problems are treated, firstly the over-saturated single intersection and secondly a pair of coupled intersections in over-saturated conditions. The aim of the control for both problems is that of minimising the aggregate delay over the length of the rush period, subject to the usual constraints on max and min green phases.

When the State equations are formulated in discrete time, the above problems can be reduced to a linear programming (LP) problem for a wide range of operational conditions. In this case, the control is acting as an area level controller, with a control interval of a few minutes. Also there was the assumption that the travel time between intersections was negligible compared to the control interval, however if this is not the case, the equations could be modified to take account of the travel time.

A generalised network is also presented, the network consists of arcs that handle time-dependent origin-destination requirements between certain pairs of nodes. Traffic streams, associated with different origin-destination pairs are viewed as different commodities, since elements of these streams are not inter-changeable. There is a travel time associated with each arc, and also a maximum flow rate (capacity). In addition there is a storage capability which is assumed to exist at the end of each arc, just in front of a node. Traffic is assumed to travel from node to node at constant speed and then be stored in front of the node until it is allowed to proceed past the node toward its destination. Service of the queues is first-in-first-out (FIFO). The final assumption is that the pre-determined paths for each unit of traffic can be roughly translated into percentage turning movements that are given functions of time. The control law is formulated, minimising the time spent while waiting in queues.

Discretisation again results in an LP problem. This paper fits into all of the three categories listed above.

Papageorgiou (1990) uses the standard State-Space model to represent the whole network problem, i.e. group three. A macroscopic modelling framework for dynamic traffic phenomena on multidestination freeway and/or road networks with time varying demands is described. The model is capable of describing traffic phenomena for different kinds of networks (e.g. freeway, networks, urban networks, and corridor systems). The model also includes a number of control measures such as traffic signals, ramp metering, real-time information and route guidance. The last two measures may be applied either in a collective way, by means of variable message signs (VMS), or individually, by direct communication with equipped vehicles. The paper describes a unified approach to the design of all the above control measures.

A very detailed model is presented. One of the most important variables, used for control via route guidance is the independent splitting rates or turning proportions. A distinction is made between global variables and destination oriented variables. Global variables do not distinguish between subflows with different destinations.

The model is built up by first defining a static model of the network nodes. Essentially, the traffic flow is distributed through the network of nodes along links according to the destination of the sub-flows and according to the splitting rates.

Next the network links are modelled. A link model is in charge of transforming the input variables/flow into output variables. There are several alternative link models which may be used for this purpose. Any model of the following standard state-space form may be used:-

$$y(k) = G\,[\,x(k), u(k)\,]$$

$$x(k+1) = F\,[\,x(k), u(k)\,]$$

where k = 0, 1, 2 ... is the discrete time index (i.e. x(k) = x(kT), T being the sample time interval. y comprises the output variables, u comprises the input variables, the state vector x depends upon the chosen model, and G, F are non-linear, differentiable vector functions.

The link is first described at a macroscopic level by the introduction of the global link variable, the traffic density (veh/km). The dynamic model of the link densities are built up from simple use of the conservation equation, i.e. you can't lose vehicles.

Papageorgiou uses these models to evaluate different methods of controlling the splitting rates through a time varying demand. A Dynamic user optimum is achieved via optimal control, which involves the minimisation of a performance index, and also by a feedback regulation.

The feedback concept has the following advantages:

i)it requires only a few calculations at each time step k. It is a real-time procedure, such that no iterations or other time consuming algorithms are required;
ii)the feedback law does not use current or future values of the process disturbances, namely the origin-destination demands of the penetration level.

The paper also considers the integration with other control measures such as traffic signals and ramp metering. The model is extended to introduce queues and flow controls for ramp metering and a similar model may be proposed for signal control.

The overall state-space model is easily extended to include a vector u which comprises all the signal/ramp control variables.

Signal optimising procedures for ramp-metering or urban traffic signals usually assume turning movements and hence route choice of drivers to be fixed. This assumption, however, does not hold true because drivers may change their routes after introduction of the new signal settings which renders the signal settings non-optimal. It is therefore desirable to perform signal optimisation taking into account user traffic assignment conditions.

An integrated control strategy design may be performed by use of non-linear optimal control theory or by application of the feedback methodology. The feedback concept appears particularly attractive for a broad class of traffic control problems, due to low computational effort, low sensitivity to unknown origin-destination demands and unknown compliance rates along with an integrated design procedure.

A similar feedback strategy to that above was tested via simulation, controlling traffic flow on the Boulevard Peripherique in Paris (Papageorgiou et al 1989). A macroscopic model was calibrated for a 6km stretch of the BP which included five on-ramps and six off-ramps; using data from 13 mainstream detector stations. The model divides the freeway into 12 sections and describes the average behaviour of vehicles in each section in terms of traffic density, mean speed and number of vehicles entering and leaving a section.

Four different ramp metering strategies were tested:-

i) fixed time control;
ii)local feedback control;
iii)co-ordinated feedback control;
iv)non-linear optimal control.

Fixed time control limits response to small disturbances, model inaccuracies etc and may lead to underutilisation or oversaturation of traffic on the mainstream.

Local feedback control aims at maintaining a desired traffic density downsteam of the on-ramps, by using occupancy and density measurements. A local control law of the following form was used:-

$$r_i(k) = r_i(k-1) - K\ [\ \rho_i(k) - \rho_i^d\ ]$$

where

$r_i(k)$ is the on-ramp volume for time step k, ramp i
$\rho_i(k)$ is the vehicle density for section i at time step k
ρ_i^d is the desired density for section i
k is the sample time index
K is the feedback gain.

This control law is an example of an Integral feedback regulator and has excellent robustness properties.

The feedback control reacts to actual traffic conditions to eliminate congestion. At the same time, underutilisation of the mainstream capacity is avoided if there is sufficient demand in the corresponding on-ramp. The main disadvantage of the feedback law is due to its local character; reaction to congestion does not occur until congestion reaches the corresponding section.

Co-ordinated feedback control:

It is possible to design multi-variable feedback control laws for linear systems using linear quadratic optimisation theory. This technique is also applicable to non-linear systems which are linearized around a desired steady-state.

Two alternative types of multivariable control are presented:-

i) a classical LQ - control law

$$\underline{r}(k) = \underline{r}^{d} - K_{LQ}\,[\,\underline{\rho}(k) - \underline{\rho}^{d}\,]$$

ii) and a linear quadratic integral (LQI) control law

$$\underline{r}\,(k) = \underline{r}\,(k-1) - K^{1}_{LQI}\,[\,\underline{\rho}(k) - \underline{\rho}(k-1)] - K^{2}_{LQI}\,[\,\underline{\rho}(k) - \underline{\rho}_{d}\,]$$

where

$\underline{r}$ is the vector of controllable on-ramp volumes.
$\underline{\rho}$ is the vector of densities
is the vector of some selected bottleneck densities

K_{LQ}, K^{1}_{LQI}. K^{2}_{LQI} are feedback gain matrices.

$\underline{r}^{d}$, $\underline{\rho}^{d}$, are desired values of the various parameters.

The two control laws are examples of multivariable state regulators, the latter with integral parts. The LQ control appears simpler than the LQI control, but it requires prespecification of desired values for all controllable on-ramps and all densities. Whereas the LQI control requires desired values for the bottleneck densities only, therefore LQI rather than LQ control is recommended for practical use on the B.P.

Due to the multivariable feedback, congestion in a given freeway section is "visible" from each controllable on-ramp. Hence LQ or LQI control may be more effecient than local control.

Non-Linear Optimal Control was also simulated so as to minimise the total travel time over a specified time horizon. The queue lengths on the on-ramps are also included in the performance index, an extra term, queues-squared may also be added for a more homogeneous distribution of on-ramp queues in space and time. The numerical solution of non-linear optimal control problems may be performed by using feasible direction algorithms using conjugate gradient techniques. Initial results for the whole of the BP were particularly promising.

In conclusion suitable use of on-ramp metering strategies improves traffic conditions compared to the no control case. In particular, the congested area is restricted in space and time, and mean speed is generally increased. As a consequence, total time spent in the system, (including total waiting time at the on-ramp queues), and total fuel consumption are decreased, the precise rate of decrease depending upon the particular conditions and control strategy applied.

Cremer and Schoof (1989) provide another study of a complete network, which fits into category three. The study considers a network containing an expressway together with the surrounding surface streets which form a corridor system. The study considers four types of control:

i) ramp metering at the entrances to the expressway;
ii)variable speed limitation imposed on individual sections of the expressway;
iii)traffic flow diversion by variable message signs at the exits of the expressway;
iv)dynamic strategies for traffic lights at the intersections of the surface streets.

Two simulation models are described, one for the expressway similar to that of Papageorgiou and one for the street sections which is basically a dispersion model similar to TRANSYT.

The traffic flow through the street network is controlled by the variable sequence of green times at each intersection. The control of the green times is non-traditional, ie. the cycle time and offset are not kept constant, and the only constraints are the upper and lower bounds on the green times.

The whole problem is tackled as an Optimal Control Problem with the aim of minimising total delay in the system using the four bounded control systems mentioned above. From an inspection of the causal structure of the problem it became apparent that the problem could be decomposed into several subproblems. Ramp metering and speed limitation influence traffic flow on the expressway while signalization at intersections influence only traffic flow on the surface street level. Route diversion, however, changes traffic conditions within both subsystems. This gave rise to a two-stage procedure for the solution of the overall problem: an upper level decision is made for route diversion optimization while at the lower level, ramp metering with speed limitation for the expressway and signalling of surface street intersections are optimized independently in parallel procedures.

The current paper is interested only in the optimization of signal settings and so only the Heuristic Optimization Search for Green times is described here.

Firstly, the control horizon or optimization period should have a reasonable length (eg. 1 hour) to get meaningful solutions. The number of control parameters is very high; if we have C signal cycles for all N intersections and two green times for each cycle, then a total of 2CN green time durations have to be determined optimally. In order to avoid high computation time a particular search strategy was developed.

First, a set of search directions have to be defined in the 2CN dimensional control space. These directions include the unit vectors, plus other vectors which increase (or decrease) green times in one direction uniformly or which shift the proportion of green times at a certain intersection for a single cycle. Then starting from a green time pattern which was optimized for steady-state conditions (eg. by TRANSYT), a random search was applied which tried to improve the traffic performance on the surface streets, given the demand and flow diversions as chosen by the route diversion decision on the upper level. The routine chooses randomly in

every search step directions from the set defined before and makes use of the knowledge about directions which were successful in previous steps.

Re-writing the simulation program to run on a vector computer, to enable parallel processing reduced the computation time by a factor of 15 for this partial optimization on the lower level.

A view of future systems is presented by Bolelli et Al (1990) in the Intelligent Intersection. The paper describes two similar approaches to the problem of Intelligent Intersection Control, which have a number of commonalities in the problem formulation and in the ultimate goals, although different paths are envisaged to the final implementation.

The first approach is that being developed by INRETS within the project "Le Carrefour Intelligent". The purpose of the project is to develop a general architecture capable of reaching the complex objectives of traffic intersection management using knowledge based techniques, computer vision, new communication devices and on-board electronics.

The second approach is that being developed by MIZAR within the project "Intelligent Intersection Controller (IIC) 2000", which combines the results of past developments in real-time traffic light control strategies (SPOT) with current developments in the integration with route guidance strategies.

Both systems aim to make use of improved knowledge of actual traffic conditions in the future, with data coming from vehicles which will give better parameters to improve the prediction mechanism for traffic control strategies.

New control actions at crucial nodes will also be available, such as route guidance, starting commands or link speed recommendations resulting in platooning actions. Furthermore they expect that some hard constraints on traffic light operation will be relaxed (eg. minimum green times), obviously without reducing safety.

The two approaches suggest a means of using all the envisaged information and control actions in an integrated manner; thus providing an optimal solution.

Yagar (1977) outlines the traditional problem of minimizing delay at a signalised intersection and then considers extending the problem to take account of transient demands through the peak.

After describing a suitable method for estimating delay with transient demands, the paper then gives an example by splitting the peak into two time slices. Each time slice is assumed to have a constant demand, the analysis is similar to the time-stationary analysis, resulting in a much larger linear programming problem. Each time slice has its own green time allocation to the various stages of the traffic signal cycle. Queueing is allowed at the end of any time slice except the last one. The demand for each time slice is equal to the new demand for that slice plus any queued demand from the previous slice. A linear programming procedure can be used to minimize the aggregated delay, including queueing, on all of the approaches to the intersection in all of the time slices of the peak period. The two time slice example was carried through as an LP problem, the first time slice of the peak period corresponds to peak demands in which there are oversaturations and queues build up. The second time slice corresponds to the subsequent period during which any queues left over from the first slice are served. The method was simulated for one intersection and gave good analytical results, however at the time of writing it was considered impossible to implement in real time over many intersections due to computer limitations. This study fits into category one, traditional minimisation

problems, and considerations such as those studied here have been developed into existing systems which are discussed in a later section.

Bacon (1977) describes how it is not always best to minimise delay and shows that it is better to maximise capacity for a chain of linked signals. In this type of approach, queues on the approaches to the signals may actually be contrived by the control system, rather than reduced to a minimum.

The paper is split into two parts, the first part considers a single link between two junctions, with the aim of obtaining a range of possible timings, or offsets, for that link, which will maintain saturation flow. An offset is defined here as the time lag between the start of the upstream green phase and that at the next junction downstream in which the same platoon crosses the stop line. This differs from the normal definition in that the offset is relative to a platoon of vehicles.

The offset must fulfil two objectives:

i)it must ensure that at any time when the lights are green at the downstream junction, there is a queue to maintain the flow; and
ii)it must ensure that the link is not too congested; ie. that there is sufficient space to accommodate vehicles leaving the upstream junction.

At this stage, a solution is usually possible, since the problem of coordinating the timings on all the various links is ignored; as are the timings in opposite directions on the same link.

The second part discusses the problem of coordination, the constraints on a solution are described; and a strategy for searching for solutions is suggested.

The type of system described here would only be successful under conditions of severe congestion.

Another traditional approach for optimising green splits is described by Macleod et al (1977). It is based on the usual minimisation of a performance index of weighted stops and weighted delays, but uses a feedback concept or "Self-Reflective Cycle" making it well suited to on-line control.

Usually, such constrained optimization is difficult, if not impossible, using conventional hill-climbing or delay comparison techniques. In addition, the commonly used approach of apportioning cycle times on the basis of the y-ratios (flow/saturation flow) of the phases has the drawback that proper account is not taken of the queues left over for the next cycle and the method is more suitable for off-line application. By using the self-reflective cycle, these difficulties are overcome. The cycle times are ultimately apportioned on the basis of some flow factors of the junctions. The approach not only makes each cycle largely independent of others, but also has the beneficial effect of stabilising the queues developed at its end during optimisation of the junction criterion function. The method is well suited to those on-line applications where there are a sufficient number of output detectors to enable accurate prediction of arrival rates to be made. The analysis still required considerable attention for the oversaturated case.

In fact there are only a few papers which tackle strategies for oversaturation (category two). Michalopoulos and Stephanopoulos (1977a, 1977b) present two in-depth theoretical studies of

oversaturated signal systems with queue length constraints, firstly for a single intersection and then for two connected intersections.

The model is based on the earlier work of Gazis et al and describes mathematically the optimal control policy minimising total intersection delay subject to the usual constraints plus the new upper bound on queue lengths. In the second study, where there are connected intersections, the coordination of intersections is also described as a constraint to the problem. It is assumed that the control action taken at the second intersection is an explicit function of the control action taken at the first intersection, with an appropriate time lag or lead.

The authors suggested that extensions of the theory to more than two intersections should not present any problem as they are trivial extensions of the two intersection case. It should be noted, however, that as the number of intersections increases or when a large number of queue length constraints are imposed, the optimal control policy becomes more complicated so that in the final analysis, the optimal strategy may not be practically applicable in a pretimed system. So as the system and performance requirements increase, closed loop control should be considered.

Clearly if an adaptive demand prediction algorithm is available, the optimal control policy could be determined and implemented in a real time basis. At the time of writing, such prediction algorithms for real time traffic control did not exist.

The case of two way streets and multiple phase operation is not a trivial extension of the theory developed here. However, if the peak hour volumes are directional, the previous theory can be applied on two way two phase networks, but the resulting control is not necessarily optimal since the peak directions will receive preferential treatment.

Some further clarifications are in order concerning the qualitative results associated with the queue length constraints. The conclusion that the signals must be switched as soon as the queues reach their upper bound seems erroneous if simple queue dynamics are considered. For example, based on physical considerations it is evident that there is a finite time delay between the switching of the signal and the time at which the last car in the queue starts moving. Thus, if the signal is switched as soon as a queue reaches its upper bound, it is certain that this upper bound will be exceeded. This result came about from the assumption that the saw-toothed patterns of the cumulative ouput curves can be neglected and replaced by straight lines. Although this assumption is fairly realistic when queue length constraints are ignored, it presents the aforementioned problem when the queues are bounded. Nevertheless, this difficulty can be taken into account by either (a) switching the signals earlier in time to allow sufficient time for the last car in the queue to start moving as soon as it joins the queue, or (b) by shifting the cumulative output curve to the right by a specified amount as suggested by Gazis.

Henry et al (1990) recognised that over-saturation not only causes a problem in the real world, but it also highlights inadequacies in current traffic models such as TRANSYT, SCOOT and PRODYN. The queue evolution model in PRODYN is a vertical queue model, similar to that of TRANSYT. This model works well for the fluid situation and for primary congestion but fails in secondary congestion because the output flow of the model is not affected by the downstream queues.

Two approaches were intended to improve the PRODYN control before and during the secondary congestion. The first one consists of the development of a model which takes into account the output limitation due to downstream queues. Theoretically this is possible, but

practically the computation time due to Dynamic Programming use and the communication between decentralized processors would increase the costs of the control system drastically. A second approach is to develop an upper level which constrains PRODYN stage times during the congested period.

The first idea was to constrain the PRODYN optimization problem with queue constraints. This is a good idea for avoidable congestion, but when unavoidable congestion appears, ie. when the capacity of the network is lower than the demand; the optimization problem has no solution and no control can be computed.

The problem was to constrain the optimization in such a way that, independently of demand, the set of solutions is never empty. In the PRODYN model, there are two variables which do not depend on traffic demand: the stage and the stage time. Constraints on the stage time were chosen because it seems related to congestion, although the relationship is non-trivial. A two-level solution was then proposed: the upper level, working with a large sampling time, analyses the demand and then sets the PRODYN stage time constraints for all intersections of the network.

This upper level seeks a compromise between two contradictory goals:

i) ensure the traffic fluidity;
ii) constrain the stage time range as little as possible.

The result of the allocation of minimum and maximum stage times during congestion is that avoidable congestion is prevented and unavoidable congestion is shared among links proportionally to their capacities. This strategy postpones the secondary congestion. The upper level works with a high sampling time which compensates for the PRODYN short sighted horizon.

Numerical experiments were run on the traffic simulation model SITRA-B using the PRODYN real time method at the lower level, although the method is quite general and may apply to other real time strategies like SCOOT, OPAC and UTOPIA. The simulation results show that the introduction of the upper level yields substantial benefits in total travel time, sometimes up to 40%. The higher the demand and the large the network, the better are the benefits.

Shepherd (1991) investigated via simulation the case of an over-saturated arterial. The strategy was based on the earlier work of Rathi (1988) which was described earlier. The aim of the strategy was to reduce the effects of blocking-back during over-saturation, whilst maintaining the throughput of the arterial near to saturation flow. This in effect sets up a form of "Automatic Gating", using information about the space downstream to set the green splits upstream, resulting in a `flared green' approach or `funnel' of green times in over-saturated periods.

The strategy was first tested using the model TRAFFICQ on the Finchley Road Study, London and later developed on a sub-network of Turin using the microsimulation model NEMIS. The Turin study showed that the method could reduce the amount of blocking-back by up to 60%, the total delay by 6% and the total travel time by 2%.

4. EXISTING TRAFFIC CONTROL SYSTEMS

This section describes some of the latest UTC systems operating or under development around the world. The systems described here are SCOOT (UK), PRODYN (France), OPAC (USA), UTOPIA (Italy), SAGE (France), STAUKO/VERON (Germany) and SCATS (Australia). This selection may be incomplete, but covers 3 continents and several different approaches to urban traffic control.

4.1 DYNAMIC CONTROL MEASURES VERON AND STAUKO

Dynamic or on-line control measures respond to the traffic situation and changes in the situation. They require a supply of traffic data, usually obtained via loops or detectors. There are many ways of treating/using the data and just as many control algorithms, but common to all methods is the control tool - the signal settings. There are three settings which are commonly used - the green split, the offset and the cycle time. One of the main objectives in on-line control is to ensure that the first vehicle to leave an upstream signal reaches the end of the downstream queue just as it moves off. This reduces the total number of shockwaves in the network.

On-line optimisation of offsets

There are a number of references by Boettger which come under this heading. They all have the same underlying objective function or performance index, namely:

$$Z = GW * W + GH * H$$

where GW and GH are weightings applied to waiting time and number of stops respectively. The number of stops is usually multiplied by half the red time to give it some time element. The general aim then in these methods is to minimise Z over an arterial road or set of junctions comprising a small network, optimising the offsets by changing the green splits. The difference in methods comes mainly in how the data is used. Note that this method is also the basis of TRANSYT. However in Britain, the technique is to try and minimise Z over an area or grid, whereas in Germany Z is optimised along major arterials.

4.1.1 VERON

In the program VERON, developed by SIEMENS, detectors are placed one per lane, 60-140m upstream of the junctions. The data is smoothed to give intensity flow profiles. These profiles depend upon the times of arrival, turning movements upstream and upstream offsets and green times. The data is collected every two seconds, the original number of vehicles is written as $A_{n,i}$ where n is the number of the cycle and i the number of the sub-interval within the cycle. $A_{n,i}$ is given one of three states 0, 1 or 2, more than two vehicles in two seconds is considered not possible. Then a smoothed value,$A_{n,i}$, is produced as follows:

$$\overline{A}_{n,i} = \overline{A}_{n-1,i} + \alpha (A_{n,i} - \overline{A}_{n-1,i}) \qquad 0 < \alpha < 1$$

where α controls the sensitivity to the new value compared to the previous cycle. Practical results show that α should be higher for increasing traffic but relatively small for decreasing traffic.

These traffic intensity profiles are formed on a cycle by cycle basis. These profiles are used to optimise, first of all one intersection for all approaches according to the previous objective function. Then between two intersections by fixing the previously optimised intersection and varying the other offset. This can then be applied to the next junction along the arterial, in fact it can be applied from the middle of the arterial outwards in both directions. It has been shown that for arterials the optimum solution is reached wherever the process is started. Each approach must be included in the optimisation path only once.

Optimisation in meshes or closed networks depends on the optimisation path, however, increasing the size of the mesh reduces the difference in performance between different optimal paths.

VERON does not change the offsets every cycle, it only makes changes in medium to extreme conditions and even then checks the improvements against a preset minimum. The program also takes account of overloads. The algorithm is modified to take account of queues persisting at the end of green. It changes the green splits to give a flared progression as mentioned earlier in the paper. This is only operated during overload as it does not optimise offsets but eases congestion. Changes in cycle time upset the algorithm for a while. If areas have different common cycle times then co-ordination between the areas can be systematic with cyclic changes.

A before and after study was carried out using VERON in Graz (SCHWARZ and PAJIC). The results were produced using the VEMA measurement processing program. The before situation was a fixed "green wave"; one point had to remain fixed to accommodate a tram-line, VERON performed well, reducing stops by 26% and waiting time by 27% over the whole route during a whole day. Fuel and exhaust emission savings were estimated to be of the same order.

Other models were put forward by Boettger, the main difference being in the data processing. The models should be simple, flexible, described by on-line measurements, they should account for platoon dispersion and the stochastic nature of traffic.

One of these models split the intensity profiles into two parabolae, one produced by straight on traffic and one by turning-in traffic. The parabolae should join continuously and negative values were not permitted. If such data was not available a second model worked just as well. It separated the platoons into straight and turning traffic but produced simple 2 block profiles for each, consisting of a main block and tail-block. These blocks were based upon maximum intensities for the main block plus a tail, important parameters were the gaps between tail blocks and front blocks.

The optimisation procedure was then the same as for VERON. However, these models did not work in overloads because the parabolae tended to join and become a straight line, representing constant intensity.

4.1.2 STAUKO

This method assumes a congested situation at critical nodes, ie. a queue remains at the end of green (Boettger, 1987). As suggested previously the green wave is not the optimum solution, therefore backward co-ordination must be attempted. Again the principle is that the first car arrives at the queue as the last car moves off. Delays and stops increase if the green wave is used, due to blocking back and the creation of more shock waves.

STAUKO first identifies critical and non-critical intersections. The method considers the flow to and from the critical node. It considers the critical intersection as fixed and controls the input and output by changing the offsets and green splits of the non-critical junction. The green ahead times may be different at the non-critical intersection so as to allow metered input and maximum output at the critical intersection. It is constrained by the fact that the green times must end at the same time so as not to affect cross-traffic at the non-critical intersection adversely. There are parameters within the algorithm which can be set to determine how far from the average offset and green time the control in each direction can be according to the queue detector data, which gives an indication of congestion in each direction. Basically the algorithm uses spare green times at the non-critical intersection via a decision matrix plan, based upon queue detector data. Intersections which are fully congested should not be used with STAUKO.

4.2 AN EXPERT SYSTEM APPROACH TO CONGESTION: SAGE

An expert system is being developed in Paris to advise and eventually replace human controllers (Foraste and Scemama, 1987). Firstly, a knowledge base has been built up consisting of facts and rules.

Facts are static or permanent; they describe the 190 junctions in the test network via links, zones and traffic signal plans. This helps the system create "congestion chains".

There are about 200 rules which are not site dependent and are therefore kept separately. The rules are a line of reasoning based on interviews with traffic experts. Examples might be the minimum green time allowed, the minimum amber time or say not to change the order of stages at a particular intersection.

The Inference Engine activates reasoning. It is data driven and forward chaining, and tries to deduce everything that can possibly happen. The rules operate on variables using first order logic, multiple instancing for each rule. Links are described by thresholds of congestion as being one of the following states: flowing, congested or unknown. The main program runs every three minutes so it can act as a control process in real time. At present it communicates to the operator through graphics and is used as a decision making aid. Its conclusions can be applied or ignored but the line of reasoning can be supplied on request.

The expert system can also interrogate the operator about video information which it cannot see. The system builds up "chains of congestion" and diagnoses the possible causes. In this way a tree-structure of congestion is built-up and the system proposes actions from a knowledge of the signal plans. The system can also generate warnings or alarm messages for the operator. A history of congestion can be generated and stored so that the system could eventually learn and predict from situations.

4.3 SCOOT

The SCOOT Urban Traffic Control System is now operational in over 40 cities in the UK and overseas (Bretherton, 1989; Bretherton and Bowen, 1991). SCOOT is a fully adaptive system which collects data from vehicle detectors and then calculates settings which reduce delay and stops. The SCOOT system has been tested and evaluated in a number of field trials using the floating car survey technique. The trials showed that on average, SCOOT reduces delay to vehicles by 12% when compared to fixed time control using up to date plans calculated by TRANSYT.

Since changes in flow pattern cause fixed time plans to age, it is estimated that SCOOT will achieve a reduction in delay of about 20% when compared to a typical fixed time system where plans are not updated yearly.

SCOOT Principles

SCOOT operates groups of adjacent junctions on a common cycle time. At any instant, the cycle time, green durations and offsets between signals are controlled by timings held in computer store. An important feature of SCOOT is the traffic model. The traffic model uses information from vehicle detectors on the approaches to each junction to predict the total delay and stops caused by the signal timings: the signal optimiser adjusts the timings to reduce this total. Frequent small alterations adapt the signals to short term fluctuations in the traffic demand. Longer term trends are satisfied by the accumulation of small alterations over several minutes. Thus, there are no large sudden alterations in timings that might disrupt traffic flow. Further, this strategy of control has a low sensitivity to detector malfunctions and avoids many of the problems of predicting traffic behaviour for several minutes into the future.

The latest release of SCOOT, Version 2.4 is described by Bretherton and Bowen (1991). Eleven new features are described in detail, the following were designed to tackle the problem of severe congestion:-

a) Gating and action at a distance

The main purpose of the "Gating" logic is to restrict the in-flow of traffic into a sensitive area to prevent the build up of long queues or congestion in that area. In order to implement "gating", SCOOT must be able to take "action at a distance"; that is, it must be able to modify the signal settings at junctions which may be far removed from the area of immediate concern.

The gating logic allows one or more links to be identified as critical, or bottleneck links. A bottleneck link can affect the green time on gated links. The gated links are those links which have been designated to store the queues which would otherwise block the bottleneck link. When the bottleneck link is too busy the green time is reduced on the gated links.

The traffic engineer plays an important role in firstly, specifying which links should be bottleneck links and which should be the associated gating links. For a bottleneck link the traffic engineer also specifies the critical degree of saturation above which he expects problems. This critical degree of saturation is used to trigger the gating marker depending on the relative sizes of the current and critical saturations on the bottleneck link. All of the logic is contained within the split optimiser. When the saturation crosses the critical saturation from either direction; ie. gating is about to start or end, then two successive stay decisions are imposed to increase the stability of the gating logic.

b) Congestion offsets

In congested conditions the desired offset on a link can be different from the one which would minimise delay on the link. The offset needs to be set so that capacity is maximised and so that the link is not blocked when the upstream junction is showing green to the critical approach. An offset may be specified which the optimisers will move the signal settings to when congestion is detected on a link. In SCOOT terms, a link is congested

when the detector is continuously occupied for at least 4 seconds; the longer the period of continuous occupancy, the higher the level of congestion. Since detectors are normally placed at the upstream end of links, congestion occurs when the link becomes blocked with traffic.

c) Congestion link facility - information from another link

The latest release makes it possible to specify that a link can use the congestion information from another link either as well or instead of its own information. Links can be specified as "suppliers" and "receivers" of congestion information, so that the signal timings can be changed accordingly.

Further research is being undertaken to use the traffic information that SCOOT provides for the formation of a traffic database, detection of incidents, monitoring of congestion and integration with a dynamic route guidance system. There is also a DRIVE II project, PRIMAVERA which aims to enhance SCOOT, producing an integrated strategy incorporating queue management, public transport priority and traffic calming techniques for over-saturated arterials. These will be developed and tested on site in Leeds by the end of 1994.

4.4 OPAC

OPAC (Optimisation Policies for Adaptive Control) (Gartner, 1989) is a computational strategy for demand-responsive decentralised traffic signal control that is being developed and tested in the United States. The strategy has the following features:

i)it calculates controls that approach the theoretical optimum;
ii)it requires on-line data from upstream detectors and from neighbouring intersections; and
iii)it forms a building block for demand-responsive decentralised control in a network.

The OPAC strategy had the following objectives:-

a)The strategy must perform better than off-line methods.
b)Development of a new control concept, better suited to the variability in traffic. The conventional notions of offset, split and cycle time, were not suited to demand-responsive control.
c)The strategy must be truly demand-responsive, ie. adapt to actual traffic conditions and not to historical or predicted values that may be far off from the actual
d)The strategy must not be arbitrarily restricted to control periods of a specified length but should be capable of frequent updating of plans, as necessary. It should be based on decentralised decision-making.

The development of the OPAC strategy proceeded in three stages. First, a Dynamic Programming (DP) procedure was developed, which served as a standard of performance for demand-responsive control, since dynamic programming is capable of generating optimal control strategies. Next the procedure was simplified to make it suitable for on-line calculation. In the third stage the procedure was further refined by introducing a Rolling Horizon approach, similar to that of UTOPIA (Mauro et Al, 1984).

The procedure uses an Optimal Sequential Constrained (OSCO) search and has the following basic features:-

i)The optimisation process is divided into sequential stages of T-seconds. A stage length is in the range of 50-100 seconds.
ii)During each stage there is at least one signal change (switchover) and at most three phase switchovers.
iii)An objective function (total delay) is evaluated sequentially for all feasible switching sequences and the sequence generating the least delay selected.

The optimal switching policies are calculated independently for each stage, in a forward sequential pass for the entire process. Computational results show that the OSCO approach provides results close to the genuine DP approach (within 10%). Thus, the stage optimisation can serve as a building-block for demand-responsive decentralised control. However, the technique requires future arrival information for the entire stage, which is difficult to obtain. To mitigate these requirements, so that only available flow data are used, the Rolling Horizon optimisation is introduced. Upstream detectors provide advance flow information for the "head" of the stage. For the "tail", data from a model is used. An optimal policy is calculated for the entire stage, but is implemented only for the "head" section. The horizon is then shifted ahead, new flow data obtained for the entire stage (head and tail) and the process repeated.

The sample time used for the rolling horizon was 5 seconds, the "head" was 15 seconds and the tail 45 seconds, giving a total horizon of 60 seconds with a "roll period" of 15 seconds.

The OPAC method was implemented and tested using the NETSIM simulation model and was also field tested in two locations (Arlington and Tuscon). Average delays were reduced by 5-15% compared with existing traffic-actuated methods, with most of the benefits occurring in high volume/capacity conditions. There is a continuing research effort in this area in the US.

4.5 UTOPIA

UTOPIA (Urban Traffic Optimisation by Integrated Automation) is the name given to the control strategies used in the real time traffic control implemented over a wide area of Turin since 1984 (Mauro, 1989; Donati, 1984).

It is an innovative hierarchical decentralised traffic light control system with the objectives of giving absolute priority to selected public vehicles and private traffic optimisation in all traffic conditions.

The first implementation of UTOPIA was over a significant area of Turin and named `Progetto Torino', and has been running successfully since 1984.

4.5.1 Basic UTOPIA concepts

UTOPIA was designed to apply to large scale systems. The global approach was:

- to decompose the whole control problem in a hierarchical decentralised way
- define proper functionals for the resulting problems, together with rules for their interaction
- define techniques and algorithms for solving these problems.

The decomposition was done topologically, a sub-problem being defined for each intersection in the controlled network.

Then a robust feedback control for the intersection and consistent rules for the interaction between intersections were found. In order to guarantee the stability at the network level, interactions were defined with an upper level too. Overall the problem was decomposed into a series of interrelated, smaller sub-problems which could be classified in two classes, the `Intersection Level' and the `Area Level'.

The Intersection Level

This is the lower level of the control scheme. At this level a `local' control operates for each traffic light intersection or zone (a zone consisting of a group of connected junctions), interacting with neighbouring zones or intersections and the area level. Every `Local Control' is further subdivided into two main parts, the `Observer' and the `Controller'.

The observer updates the estimate of the `State' of the intersection based on the available data, such as traffic counts from loop detectors and traffic light states. The observer uses a microscopic model of the intersection, where the state elements are the number of vehicles to be served for every incoming link, grouped in steps of 3 seconds. The state vector for an intersection is the composition of the state vectors related to the incoming links. The state vector for an incoming link is derived from the vehicles already in the link, ordered and grouped by predicted arrival times at the stop line. The observer produces estimates such as queue lengths, travel times per link, turning percentages and saturation flows.

The controller determines the signal settings to be applied to the traffic lights. It optimises a suitable functional (described later), specific to the current traffic situation around the intersection. Optimisation is over a `Time Horizon' consisting of the next 120 seconds, and is repeated every 3 seconds, the resulting optimal signal settings are actually in operation for the next 3 seconds after which the process is repeated. The so obtained closed loop control can be viewed as an `Open Loop Feedback Control' or as an application of a `Rolling Horizon' concept.

In order to guarantee the optimal control at the network level, the functional to be optimised was designed with a strong interaction concept, taking into account the state of neighbouring intersections giving a closed loop capability of building dynamic signal coordination. The local controller is also constrained by limits given by the area level control, such as maximum and minimum stage lengths, or order of stages.
In more detail, the functional to be optimised is defined by the sum of many weighted costs such as time lost by vehicles, number of stops, queues, time lost by public vehicles, deviation from the signal setting decided in the previous iteration. These are broad descriptions of some of the cost elements presently used, and they are currently being updated to take account of over-saturation.

The Area Level

The area level control function is also split into two main modules, the `Observer' and the `Controller'.

The observer analyses the traffic conditions over the whole area, based on actual traffic counts from the network and filtered statistical traffic characteristics. It predicts, in real time, the main routes which will be taken by private traffic and the flow levels (Demand) at the origins of the predicted routes. The observer is based on a discrete time model with a time interval of

3 minutes. The network of major routes is represented by a macroscopic model consisting of `Storage Units'. Storage units correspond to pieces of an arterial or to intersections (nodes). A series of fixed routes is superimposed on the network. The state variables are defined as the number of vehicles in each storage unit per route per time step.

The controller optimises the network functional by acting on controls: average speed and saturation flows within each storage unit. The network functional represents the `total travel time' spent by private vehicles in crossing the area. Once again it is optimised by OLFC (or rolling horizon) techniques over a horizon of 30 minutes.

4.5.2 The dynamic traffic model

The urban traffic is described by three models:-

- the macroscopic model of private traffic, representing the behaviour of traffic streams over the whole controlled area
- the microscopic model of private traffic, giving a detailed representation of the private traffic at each intersection
- the public traffic model which describes in detail the behaviour of every vehicle for the lines with priority.

The macroscopic model of private traffic

The major directions of traffic are represented as inter-connected "Storage Units". Several intersections with traffic lights can form one storage unit. Two parameters are used to describe the vehicle law of motion:

α^k is related to the average speed on unit k
β^k is related to the saturation flow on unit k.

Then calling x_i^k the number of vehicles on unit k during the i^{th} interval, and O_i^k the number of vehicles which can leave the same unit k during the i^{th} interval, the model assumes:-

$$f^k\ (X_i^k,\ \alpha_i^k,\ \beta_i^k)$$

The macroscopic model is used for:-

- recognition of traffic perturbations at their origin, performed in real time, through data collection from detectors.
- continuous identification of parameters characterising the traffic conditions.
- and finally the model is used for the area level control algorithms.

The microscopic model of private traffic

The microscopic model describes every intersection with traffic lights as a discrete dynamic system. An intersection is regarded as a set of links. The "State" of an intersection is given by the vector of arrivals at the intersection itself, for those vehicles already on an approaching link:-

$$(n_i^{k,1}, n_{i+1}^{k,1} \ldots\ldots, n_{i+L}^{k,j} \ldots\ldots,)$$

where:

$n_{i+L}^{k\,j}$ represents the number of vehicles already on link j, of intersection k, and at a distance from the intersection equal to L time intervals. (present time is i).

The propagation law is deterministic and is a known function of the traffic lights state at the intersection:-

$$= f\ (y_i^k\ ,\ c_i^k\ ,\ y_i^m\ ,\ c_i^m\ ,\ y_i^n \ldots\ldots)$$

where:

c_i^k is a vector representing the state of the traffic lights at intersection k at time i.

$y_i^m\ ,c_i^m\ ,y_i^n$ etc. refer to intersections m, n etc. adjacent to intersection k.

This model depends upon some parameters (typically: "percentage turns", "average overall travel speed" and "saturation flow".

The model is used for:

- local control
- continuous identification of the above mentioned parameters
- diagnosis of anomalous traffic or plant conditions.

The public traffic model

The departure time and route of each vehicle is assumed to be known and enforced. It is assumed that the travel time in each section can be subdivided into three parts:

- free travel time
- waiting time at stop or station
- lost time at intersections with traffic lights.

The third component is subject to the system control, whereas it is assumed that the first two can be described by a stochastic process.

The model is used on-line:

- to forecast, through bidimensional linear filtering, the individual time components of each section
- to diagnose automatically any disturbances.

4.5.3 The control

The Control aims to minimise the total time lost to private vehicles, subject to the constraint that public vehicles with priority shall not be stopped at traffic lights.

The control system follows the general scheme already described with two hierarchical level, one at the area level, the other at the local level.

The area level control

The goal of the area level control is to optimise overall performance via a performance function P viz:

$$\sum_i \sum_k P^k(\ x_i^k \ , \alpha_i^k \ , \beta_i^k \ )$$

where x_i^k , α_i^k , β_i^k have been described earlier.

The cost P^k associated with each storage unit k is usually linear. However in some cases it may be useful to use quadratic terms in P. The sum with respect to time (i) is over an interval for which comparatively sure information can be obtained (from half an hour to one hour).

The area control transforms the results α^k and β^k into suitable reference rules for the local controllers, roughly speaking into a "Reference Plan" and weights to be assigned to individual components of local cost functionals.

The local control

At every decision instant, the local controller k, minimises the following goal:-

$$\sum_i l^k(\ y_i^k \ , c_i^k \ )$$

where the sum is extended over 2 minutes, with a sample time of 3 seconds.

The functionals l^k take into account weights, dynamically assigned by the area level controller, to the following:-

- time lost by vehicles at intersection k
- the number of vehicle stops at intersection k
- the maximum queues at intersection k
- the same parameters, but from the adjacent intersections as caused by the propagation after intersection k
- the correspondence of the decisions on state c^k with those given by the reference plan, to be able to change dynamically, the interaction from the area level.

The public vehicles with priority are seen as constraints for the optimisation problem. The system has been running in Turin since 1984 and is constantly being updated. The benefits due to implementation of the system have been recorded as an increase of 15% in average speed for private vehicles and 28% for public transport with priority.

The latest version of the local controller called SPOT is also being enhanced in the DRIVE II project PRIMAVERA. Field trials will be carried out in Leeds and Turin by the end of 1994.

4.6 PRODYN

PRODYN is a real time traffic control algorithm developed over the last decade by CERT (Henry and Farges, 1989). It has been implemented and tested on the Zone Expérimentale et Laboratoire de Trafic de Toulouse (ZELT) system (Henry, 1989).

The main characteristics of the method are:

i)a five second sample time. The control is the decision whether to switch-over from one stage to another stage;
ii)the use of two inductive loop sensors per lane: one at the entrance of the link, the other 50m from the stop line;
iii)explicit minimisation of total delay;
iv)use of automatic control methods: Bayesian estimation, Dynamic Programming and decentralised methods.

4.6.1 Control strategy

Control

The control is the decision to switch-over from one stage to another stage.

States

The PRODYN method is based on the description of traffic behaviour by a set of discrete-time non-linear state equations. The states are:-

i) the stage and time elapsed since the last switch-over for each intersection;
ii)a variable times the saturation flow which models the non-priority movements, jamming and the vertical queue for each link;
iii)for each five seconds, the number of vehicles travelling at free speed on each link.

Constraints

Constraints on the control are authorised stage switch-overs only and maximum and minimum stage durations.

Criterion

The criterion used is the total delay, which is given by the sum of the vertical queues over all links and sampling times. It is approximated by the sum over all queues over sample times belonging to the rolling horizon plus a function depending on queues and stage at the end of the horizon.

Rolling horizon

The control computed by PRODYN is the result of a rolling horizon strategy; it is an Open Loop Optimal Feedback. The control strategy is to optimise over the whole horizon, taking into account the predicted queues and arrivals. The control is then implemented for the next sampling period and the process is repeated.

Coordination

The PRODYN controlled intersections co-operate by sending information to their neighbours about the outputs resulting from the application of the optimal policy.

This information is used to forecast arrivals for the first part of the rolling horizon; the later part of the horizon is predicted by averaging the last measurements at the entrance of the link.

State estimation

All state variables, except queues are predicted on an open loop basis (simple integration of the state equation). For each queue a discrete state estimator/predictor is used.

Optimisation

The intersection optimisation is performed via a Forward Dynamic Programming (FDP) algorithm, which differs from a traditional one in the following ways.

i)Instead of computing cost functions at grid points using the reverse state equation, it performs forward comparisons in a state space subset.
ii)Memory is not allocated to all existing subsets but only to those subsets which are effectively reached during the optimisation.
iii)The elapsed time since the last switch-over is not treated as a state.

The experiments on the ZELT system showed average gains in total travel time of 10% with a 99.99% significance.

PRODYN is also capable of real time estimation of traffic parameters such as turning movement ratios and saturation flow rates (Kessaci et al 1989).

An `Upper Level for Real Time Urban Traffic Control' was developed to deal with congested situations (Kessaci et al 1990). It was recognised that PRODYN's vertical queue model was valid only for under-saturated conditions. Indeed it didn't take into account output limitations due to downstream congestion. When this situation occurs it is useful to constrain states and/or controls of the optimisation problem to force the system to stay in a state where the model is valid. The upper level control determines the minimum and maximum network junction stage times.

The level compromises two contradictory goals:

i) to ensure traffic fluidity;
ii)to relax state constraints.

The compromise was obtained via a Gauss-like optimisation method and simulated using SITRA-B and PRODYN at the lower level of control. Simulation results showed that the introduction of the upper level yields substantial benefits for total travel time. The higher the demand and the larger the network the higher are the benefits.

4.7SCATS

The Sydney Co-ordinated Adaptive Traffic System, or SCATS was first introduced into Sydney in 1964 and computerised in 1972 (Sims and Finlay, 1984).

The SCATS adaptive method requires a predetermined data base inclusive of a library of phase split plans and offset plans for each intersection, and offset plans and control parameters for each sub-system. The ultimate performance of SCATS depends upon the accuracy of this data base in terms of catering for the expected range of traffic conditions and in conforming to the requirements of the SCATS software.

The aim of SCATS is to minimise dynamically, by plan selection, a performance index identical to that of TRANSYT.

SCATS controls over 1200 sets of traffic signals in Sydney, over 150 of which are in the Central Business District (CBD). Other systems have been based on SCATS, such as SCRAM in Melbourne and PACTS in Perth. The Australasian wide use of the system will result in 49 SCAT systems controlling 4400 traffic signal sets.

Studies on a typical urban arterial road have shown that in comparison with unco-ordinated operation, SCATS produced the following benefits:-

- Travel time was reduced by 23%
- Vehicle stops were reduced by 46%
- Accidents were reduced by 20% and
- Fuel consumption was reduced by 12%.

5. CONCLUSIONS AND RECOMMENDATIONS

First of all the objectives and constraints of the present SERC funded project must be re-iterated. The objectives were stated in the case for support as follows:-

i)To review current procedures for the `gating' of traffic in signalised corridors and networks.
ii)To develop the application of state-space control theory to the formulation of gating strategies.
iii)To enhance the strategies developed in (ii) through field trials.

Firstly, this review was not restricted to `gating' but aimed to give an overall view of current practice, especially in the treatment of over-saturation.

Secondly, the project's main aim is to implement strategies in field trials. This can be achieved by collaboration with the DRIVE II project PRIMAVERA, which aims to develop an integrated approach to queue management, traffic calming and public transport priority measures on in-bound arterials. PRIMAVERA will conduct two field trials, one on the Dewsbury Road in Leeds and the other on a corridor in Turin. The traffic signals will be controlled by the Italian system called SPOT in Turin and by SCOOT and SPOT in Leeds. These systems are therefore a constraint to the current project, in that there is no time allocation for re-writing a complete UTC system. The internal model used by SCOOT and SPOT will therefore have to be assumed to be correct for field trial purposes.

However, it was shown in the review that the definition of the model is very important when dealing with congested periods and in fact some vertical queue models would be inappropriate. It was also shown that, when dealing with over-saturation, the control or strategy should be able to see or act over a long time horizon, ie. there should be some form of upper level or area control, perhaps as well as a local immediate control. The two levels would require different traffic models, one detailed looking over a short time-horizon, and one less detailed with a much longer time-horizon.

Another of the objectives of the project was to formulate the whole problem in state space control theory, it is therefore suggested that two paths be followed in the research. The first path will be purely theoretical, the second will be steered towards field trials. Both paths of research will use the microsimulation model NEMIS as the `real world' model. This model will

be used in the PRIMAVERA project and will be set-up to model both field trial sites. It will provide loop detector data in one direction and actuate the desired signal control in the other direction.

The first theoretical path will use NEMIS as the real world throughout; it will then investigate various types of traffic model using detector data. Two types of model are suggested:-

i)density based model, similar to that of Papageorgiou et al developed for a freeway. This may be micro or macro depending on the time horizon required;
ii)a neural network based model, this would be achieved through collaboration with another SERC funded project. It would provide an innovative model if successful, the approach would be to supply various sets of data in the first instance and if successful it could be developed further.

Once the model has been tested and defined, the control action or strategy may be specified. Again various control strategies may be tested but two are put forward here for consideration:-

i)Linear quadratic optimal control based on some mix of relevant performance indices.
ii)Regulation to desired densities.

Both methods may be formulated in state-space and programmed within the NEMIS model.

The second path also uses NEMIS as the real world and is the same path which will be used in the PRIMAVERA project. NEMIS will be able to interface directly with both SPOT and SCOOT, again sending detector data in one direction and receiving the control in the other direction. In this case, any strategies will be developed directly within the UTC systems, or as add on modules for SCOOT. Clearly a knowledge of the SPOT cost functions is required and any strategy will have to be written in terms of these cost functions for implementation in the SPOT system. For the SCOOT system, the kernel is not available and the strategy will be written as an add on module with its own objectives.

Both systems will be tested via simulation prior to installation and field trials. It may be that following this path will prevent a theoretical formulation of the problem and indeed may result in an empirically developed algorithm. If this turns out to be the case then it may be useful to compare the results of path one with those of path two to find where improvements may be made.

6. REFERENCES

BACON, W (1977). Linked traffic signals for maximum capacity. *Transportation Research*, **11**, pp 183-188

BOETTGER, R - SIEMENS. The structure of urban traffic control and its technical realization.

BOETTGER, R - SIEMENS (1987). Koordinierung von Signalanlagen in Stausituationen (STAUKO).

BOETTGER, R (1968). Simulation of road traffic with data processing systems. *Neue Technik* **A4**, pp 213-224

BOETTGER, R - SIEMENS (1971). Optimal coordination of traffic signals in street networks. *Fifth International Symposium on the Theory of Traffic Flow and Transportation*, held at University of California, Berkeley, June 1971.

BOETTGER, R (1982). On-line optimisation of the offset in signalised street networks. *IEE International Conference. Road Traffic Signalling, IEE Conference Publication No* **207**

BOLELLI, A, SELLAM, S and SEIDOWSKY, R (1990). Intelligent intersection. *IEE Conference Publication No* **320**, pp 81-90

BRETHERTON, RD (1989). SCOOT urban traffic control system - philosophy and evaluation. *IFAC Symposium on Control, Communications in Transportation,* September 1989, pp 237-239

BRETHERTON, RD and BOWEN, GT (1990). Recent enhancements to SCOOT - SCOOT version 2.4. *3rd International Conference on Road Traffic Control. IEE Conference Publication No* **320**, pp 95-98

CREMER, M and SCHOOF, S (1989). on control strategies for urban traffic corridors. *IFAC Control, Computers, Communications in Transportation,* Paris, France, pp 213-219

Communications in Transportation, 6th IFAC/IFIP/IFORS Symposium on Transportation, held in Paris, September 1989, pp 253-255

D'ANS, GC and GAZIS, DC (1976). Optimal control of over-saturated store-and-forward transportation networks. *Transportation Science*, **10**, pp 1-19.

FORASTE and SCEMAMA - INRETS (1987). An expert system approach to congestion.

GARTNER, NH (1989) OPAC: Strategy for demand-responsive decentralized traffic signal control. *IFAC Control, Computers, Communications in Transportation,* Paris, France 1989, pp 241-244

GAZIS, DC and POTTS, RB (1965). The oversaturated intersection. *Proc.2nd International Symposium on the Theory of Traffic Flow*, held in London 1963, Paris OECD.

HENRY, JJ (1989). PRODYN tests and future experiments on ZELT. *VNIS '89: Vehicule Naviagation and Information Systems, IEEE Conference,* held in Toronto, September 1989.

HENRY, JJ and FARGES, JL (1989). PRODYN. *CCCT '89: Control, Computers,* RATHI, AK (1991) Traffic metering: an effectiveness study. *Transportation Science*, July 1991, pp 421-440

HENRY, JJ and FARGES, JL (1989). Traffic congestion control. CCCT '89: Control, Computers, Communications in Transportation. *6th IFAC/IFIP/IFORS Symposium on Transportation* held in Paris, September 1989.

HUDDART, KW and WRIGHT, C (1989). Catastrophic traffic congestion and some possible ways of preventing it. *Proc. TRAFFEX International Traffic Engineering Exhibition. Seminar on Congestion, Control and Parking Enforcement*, held in Brighton, April 1989.

HUDDART, KW and WRIGHT, C (1989). Strategies for urban traffic control. The Rees Jeffreys Road Fund "*Transport and Society*".

HUNTER, G and OLDRIDGE, B (1989). A new concept for pricing congestion.

KESSAEI, A, HENRY, JJ and FARGES, JL (1990). Upper level for real time urban traffic control systems. *11th World Congress of IFAC,* August 1990.

MAY, AD, MONTGOMERY, FO and QUINN, DJ (1988). Control of congestion in highly congested networks. *Proc. CODATU IV Conference,* held in Jakarta, June 1988.

MAURO, V and DI TARANTO, C (1989). UTOPIA - *CCCT '89 - AFCET Proceedings* September 1989 - Paris, France.

MAURO, V, DONATI, F, RONCOLINI, G and VALLAURI, M (1984). A hierarchical decentralized traffic light control system, the first realization. *IFAC 9th World Congress,* Vol.II, 11G/A-1. 2853-58.

MICHALOPOULOS, PG and STEPHANOPOULOS, G (1977a). Over-saturated signal systems with queue length contraints - I. Single intersection. *Transportation Research*, **11**, pp 413-421.

MICHALOPOULOS, PG and STEPHANOPOULOS, G (1977b). Over-saturated signal systems with queue length constraints - II. Systems of intersections. *Transportation Research*, **11**, pp 423-428.

PAPAGEORGIOU, M (1990). Dynamic modelling, assignment, and route guidance in traffic networks. *Transportation Research* **24B**, 6, pp 471-495.

PAPAGEORGIOU, M, HADJ-SALEM, H, BLOSSEVILLE, JM and BHOURI, N (1989). Modelling and real-time control of traffic flow on the boulevard peripherique in Paris. *IFAC Control, Computers, Communications in Transportation,* Paris, pp 205-211.

PRIMAVERA: priority management for vehicle efficiency environment and road safety on arterials. *DRIVE II Project V2016.* Commission of the European Communities DGXII, Telecommunications, Information Industries and Innovation.

RATHI, AK (1988). A control scheme for high density traffic sectors. *Transportation Research*, **22B**(2), pp 81-101.

SCHWARZ, H and PAJIC, S - SIEMENS. Adaptive offset optimisation (VERON). *In*: Graz, Grünlickt Nº26, pp 6-11.

SHEIKHOLESLAM, S and DESOER, C (1989). Longitudinal control of a platoon of vehicles. Department of Electrical Engineering and Computer Science, University of California, Berkeley.

SHEPHERD, SP (1991). The development of a real-time control strategy to reduce blocking-back during oversaturation using the microsimulation model NEMIS. *Proc. 24th ISATA International Symposium on Automotive Technology and Automation,* held in Florence, Italy, 20th-24th May 1991, pp 151-158.

SIEMENS Traffic Control Computer VSR 16000 M/R. Adaptive offset optimisation in signal controlled road networks. Control Program VERON. *Traffic Engineering.*

SIMS, AG and FINLAY, AB (1984). SCATS. Splits and offsets simplified (SOS). *ARRB Proceedings* **12**, Part 4, 1984.

SINHA, PK (1977). Non-interacting control of a string of moving vehicles. *Transportation Research*, **11**, pp 109-116.

SMITH, MJ (1988). Optimum network control using traffic signals. *In Proc. Colloquium on UK Developments in Road Traffic Signalling, Institution of Electrical Engineers,* held in London, May 1988.

TOMEZUKA, M and PENG, H (1989). Vehicle lateral control for highway automation. Department of Mechanical Engineering, University of California, Berkeley.

WIE, BW (1989). Dynamic network equilibrium traffic assignment and control theoretic approach. *Paper presented at 5th World Conference on Transport Research,* held in Yokohama.

WEI-BIN ZHANG and PARSONS, R (1989). *An Intelligent Roadway Reference System for Vehicle Lateral Guidance/Control.* Institute of Transport Studies, University of California, Berkeley.

YAGAR, S (1977). Minimizing delays for transient demands with application to signalized road junctions. *Transportation Research*, **11**, pp 53-62.

wptnlist\wp349.sps