Transportation Research Part B 177 (2023) 102814

Contents lists available at ScienceDirect

Transportation Research Part B

journal homepage: www.elsevier.com/locate/trb

With spatial queueing, the P_0 responsive traffic signal control policy may fail to maximise network capacity even if queue storage capacities are very large

Michael J Smith [a], Francesco Viti [b], Wei Huang [c], Richard Mounce [d,*]

[a] Department of Mathematics, University of York, England
[b] Department of Engineering, University of Luxembourg, Luxembourg
[c] School of Intelligent Systems Engineering, Sun Yat-sen University, China
[d] University of Aberdeen, Scotland

ARTICLE INFO

Keywords:
Capacity-maximising traffic control
Quasi-dynamic traffic assignment
Responsive traffic control

ABSTRACT

The local responsive traffic signal control policy P_0 was designed to maximise network capacity under certain conditions and it has been shown, in Smith (1979a, b, 1980) and Smith et al. (2019a, 2022), that the P_0 policy and related policies do indeed maximise the capacity of many steady state networks or quasi-dynamic networks with vertical and spatial queues under various conditions. This current paper shows, by giving an example, that if queueing is spatial then the original policy *P0* itself *may not* maximise network capacity, even if the queue storage capacity of each link is very large.

Introduction

Traffic control on road networks has been a transportation research subject since the seminal work of Webster (1958), and the complexity of dealing with the interaction between traffic control and route choices has been recognised since the works of Allsop (1974), Gartner (1977) and Smith (1979a). Still, after more than 60 years, finding a general traffic control policy capable of automatically maximising network capacity while accounting for the response of the road drivers is yet to be found when spatial and temporal dynamics of queues and delays are involved. In this paper, we show, via a simple example, how one local control policy which has been demonstrated to maximise network capacity under the assumption of a point-queue process at the traffic control signals, does not always maximise network capacity when the spatial extent of queues is included – even if the queue storage capacities are very large.

This paper utilises an equilibrium traffic assignment model initially introduced by Thompson and Payne (1975). They formulated a rigid (or inelastic) demand steady state equilibrium model with capacity constraints and considered vertical queueing delays as independent variables (not given by a cost flow function); they identified these vertical queueing delays as Lagrange multipliers associated with the capacity constraints in an optimisation model. This model may we suggest be called a quasi-dynamic model because the model has explicit (vertical) queues; it is not a fully dynamic model as (i) demand is constant and (ii) only equilibria are studied. A fairly full statement of this model, including green-times, is given in Smith et al. (2019a).

Many researchers have sought to model capacity-constrained traffic assignment, aiming to achieve a more realistic or tractable

* Corresponding author.
E-mail address: r.mounce@abdn.ac.uk (R. Mounce).

https://doi.org/10.1016/j.trb.2023.102814
Received 1 August 2022; Received in revised form 21 June 2023; Accepted 11 August 2023
Available online 9 September 2023
0191-2615/

model with a better representation of the actual length or spatial extent of queues. Queue storage capacities and queue spillback still require further attention within fully dynamic, quasi-dynamic and steady state traffic assignment (especially if green-times are involved), despite work done on this by, for example, Lam and Zhang (2000), Gentile et al. (2007), Zhang et al. (2013), Bliemer et al. (2014), Bliemer and Raadsen (2017), Haddad and Zheng (2018), and also many others.

Adding control

This Section 1.1 and the Section 1.2 below show how certain traffic control policies may be added to the Thompson and Payne (1975) assignment model with queues. The control policies here use signal stages and stage pressures; the control policies we consider in this paper are all "pressure driven" where green-times swap away from stages with a lower pressure toward stages with a higher pressure, at each junction. The aim is to equalise the pressures on all stages at each junction; but constraints on stage green-times must be respected.

We consider two traffic control models: an on-and-off model (in 1.2.1 below) and a continuum model (in 1.2.2 below). Han et al. (2014) compare these two models and Han and Gayah (2015) consider further and also extend the continuum signal model. In the main body of the paper we utilise the continuum signal model.

In the on-and-off control model, at each signal-controlled junction: at each time *t*, a traffic signal gives green to a set of lanes approaching the junction, while all other approach lanes at the same junction are shown red. A single period of time during which this set of lanes is given green is in this paper called a "stage". As time passes the signal switches green from one set of lanes to another. All-red is permitted: in this case the set of lanes given green is empty. In our modelling here every separate real-life single carriageway lane is represented by a "link" in a network. The simplest of real-life responsive traffic control systems have fixed cycle times and a set of fixed sets of non-conflicting approaches at each junction; where these sets of non-conflicting approaches must all be shown green in the same order in each cycle.

At each junction the durations of the stage green-times in each cycle must add to the cycle time and also there are minimum green-times for stage durations so as to allow traffic flowing during one stage to clear the junction before conflicting traffic flows are given green during a subsequent stage. Van Vuren and Van Vliet (1992) and Smith and van Vuren (1993) consider traffic equilibrium with responsive traffic control, Smith (2010) shows how responsive P_0 may be utilised to design fixed time signal timings. Taale and van Zuylen (2001) review the traffic control and route choice problem and proposed solution methods.

Pressure-driven traffic signal control in more detail

Compatible sets of approaches to a junction and an on-and-off control model

Suppose given a capacitated network [*N, L*], with a set of nodes *N* and a set of links *L*. In this paper we suppose that each link represents a single traffic lane with a single exit saturation flow.

We also suppose that at each node (representing a junction in real life) there are given sets A_j of approach links which may be shown green simultaneously; all traffic permitted on the approach links in each A_j is here to be non-conflicting. Such a set of non-conflicting approaches will here be called a *compatible set of approach links.*

In the on-and-off control model we suppose that at each node there is a given (ordered) sequence

$$[A_1, A_2, \ldots, A_n]$$

of compatible sets of approaches. In each cycle these are given green during *n* contiguous time-periods called stages:

$$[Stage_1, Stage_2, \ldots, Stage_n].$$

We suppose that the cycle time here is to be fixed at α seconds and that for each *j* the set A_j of compatible approach links is shown green during the time period $Stage_j$ of duration αG_j seconds. We suppose that the ordered sequence

$$[\alpha G_1, \alpha G_2, \ \alpha G_3, \ldots \alpha G_n]$$

of stage green-time durations is constrained by a set of linear inequalities, in part to ensure that each link has a minimum green-time in each cycle and also to ensure that the sum of the αG_j add to α (the cycle time). See Allsop (1971) for a picture of such constraints.

In the on-and-off model; at each junction the durations of each stage during each cycle may be determined at the start of that cycle, using the flows, queues and green-times previously experienced, in the previous cycle. In this paper we consider such systems where stage green-times only move from one cycle to the next by swapping from less-pressurised stages to more pressurised stages. The pressure on stage *j* may be thought of as the stress that stage *j* is under; it then makes sense to allow green-time swaps away from stages with a low pressure (or stress) toward stages with a higher pressure (or stress). Such a responsive control system is said to be "pressure-driven"; examples of stage pressures are given below in Section 1.3.

Compatible sets of approaches to a junction and a continuum model of control

In the continuum signal control model in this paper we suppose that at each node there is a given set

$$[A_1, A_2, \ldots, A_n]$$

of compatible sets of approaches. We also suppose, for simplicity in this paper, that each link i belongs to *exactly one* A_j. Now, in the continuum model, at each time t, A_j is given green for a green-time proportion $G_j(t)$ where: at each time t:

1 the green-time proportions $G_j(t)$ add over j to 1 and also
2 the set $\{G_j(t)\}$ satisfies a set of linear inequalities, partly to ensure that each link has a minimum green-time.

We again suppose here that flows exiting from different links in the same A_j do not conflict or interact. In this paper we will refer to each A_j as a "stage" in the continuum model. We let:

$Q_i(t)$ = queue size (number of vehicles) in a queue at the exit of link i at time t (vehicles),
s_i = saturation flow at the link i exit (vehicles/second),
$b_i(t)$ = bottleneck delay felt by traffic exiting link i at time t (seconds),
$p_{iu}(t)$ = proportion of flow exiting link i at time t which then traverses signal-controlled link u,
$x_i(t)$ = flow out of the link i exit at time t (vehicles/second), and
$G_j(t)$ = proportion of time that stage A_j is green at time t.

Some stage pressure formulae and pressure-driven control policies

Probably the earliest example of a pressure-driven traffic signal control policy is the equisaturation policy specified by Webster (1958). In this policy the pressure on stage j equals the maximum of the degrees of saturation $x_i/(s_iG_j)$ (using the notation in Section 1.2.2 above) over the links i given green during stage j; and Webster proposed that stage green-times should be chosen so that all stage pressures at a single signal-controlled junction are equal, or as equal as possible. Webster regarded this simple rule as a reasonable approximation to the policy of minimising the total rate of delay experienced at the junction, assuming that traffic flows are fixed, unaltered if the signal timings are changed. Webster suggested this equisaturation policy as a way of selecting stage green-times for implementation over a fairly long period of time, the morning peak or the inter-peak for example. So his equisaturation rule for setting signals was aimed at choosing fixed time signal settings; however "equisaturation" is often utilised as part of modern dynamic traffic signal control systems.

Consider a single junction and let

$$A(j) = A_j \text{ for } j = 1, 2, 3, \ldots, n.$$

Smith (1979a, 1980) introduces the stage j pressures:

$$Press_j = \sum_{link\ i\ belongs\ to\ A(j)} s_i b_i.$$

These stage j pressures lead to the P_0 control policy which seeks to equalize these pressures over the stages at each junction. A dynamical version of this policy (stated in Smith et al. (2022) and elsewhere) swaps green-time from less pressurised stages to more pressurised stages at each junction, aiming to equalise stage pressures at each junction.

Smith et al. (2019a) introduce, for a constant vector $\mathbf{h}$, policy $P_{\mathbf{h}}$, a biased form of policy P_0. This policy utilizes the stage j pressures

$$Press_j = \sum_{link\ i\ belongs\ to\ A(j)} (s_i b_i + h_i),$$

and again policy $P_{\mathbf{h}}$ swaps green-time from less to more pressurised stages aiming to equalise the stage pressures at each junction. Smith et al. (2019a) also introduces, for a vector $\mathbf{k}$, policy $P_{\mathbf{hk}}$, a biased spatial form of policy P_0. This policy utilizes the stage pressures

$$Press_j = \sum_{link\ i\ belongs\ to\ A(j)} (k_i s_i b_i + h_i),$$

and again policy $P_{\mathbf{hk}}$ swaps green-time from less to more pressurised stages aiming to equalise the stage pressures at each junction.

Varaiya (2013a, b) and Wongpiromsarn et al. (2012) introduced the stage pressures

$$Press_j = \sum_{link\ i\ belongs\ to\ A(j)} \left(s_i \left(Q_i - \sum_u p_{iu} Q_u \right) \right).$$

The corresponding control policy is called the MaxPressure policy or *MP*. This swaps green-time away from less pressurised stages and toward more pressurised stages aiming to equalise all stage pressures at each junction. The idea of using backpressure terms, like $-\sum_u p_{iu} Q_u$ above, is suggested in telecommunication networks by Tassiulas and Ephremides (1992). This is an attractive idea for traffic control (suggested in this context by Varaiya (2013a, b) and Wongpiromsarn et al. (2012)) since it permits a pressure-driven control policy at a junction J to take some account of downstream congestion when updating the signal timings at junction J. Levin et al. (2020) extends MaxPressure to take account of the cyclical phase structure of traffic signals.

Contribution of this paper

It has been shown in past papers that policy P_0 maximises network capacity in many networks. *This paper shows that the P_0 policy does not always maximise network capacity when queueing is spatial, even if link queueing capacities are very large.*

A brief summary of past P_0 studies

Smith (1979a, b, 1980) show that, within a steady state model with cost-flow functions involving green-times, an equilibrium consistent with the original P_0 policy exists if demand is within network capacity. (We say that P_0 maximises the capacity of the network.)

Smith (1987) and Smith et al. (2022) extend this result to the Payne-Thompson quasi-dynamic equilibrium model, which has explicit queues, provided queueing is vertical. (We say that P_0 maximises the capacity of the network.)

Smith et al. (2019a, 2022) extend the above capacity-maximising results to take some partial account of spatial queueing rather than vertical queueing. In order to achieve these extensions two new policies are introduced: a biased form of policy P_0 called $P_{\mathbf{h}}$ and a biased spatial form of P_0 called $P_{\mathbf{hk}}$. Here the bias vector $\mathbf{h}$ represents the estimated attractiveness of alternative routes, and helps $P_{\mathbf{hk}}$ to reduce unnecessary queueing delays and $\mathbf{k}$ represents the degree to which links are full of vehicles and helps $P_{\mathbf{hk}}$ to reduce link queues and overflows. These queues are however, although reduced, typically not eliminated and overflows are typically not eliminated by any of these P_0 variations if pricing is not involved. Smith et al. (2022) extends these ideas to include prices (as well as delays) and shows that a pricing variant of P_0 does eliminate queues at equilibrium while maximising network capacity.

Smith et al. (2019b) shows that to maximise throughput in some simple networks when demand grows to exceed the network capacity the responsive signals (following P_0) should be "frozen"; yielding fixed time, unresponsive, signal timings as soon as demand exceeds network capacity.

The results in the P_0 papers above are all essentially positive, even if the results are only very partial and depend on assumptions which do not always hold in reality. The P_0, $P_{\mathbf{h}}$ and $P_{\mathbf{hk}}$ capacity-maximising results above depend on the particular quasi-dynamic model utilised and this is described in detail in Smith et al. (2019a).

This paper, on the other hand, gives a negative result: the paper shows that if queueing is spatial and the routeing follows Wardrop equilibrium (with all travellers on quickest routes) then the P_0 responsive control policy does not always maximise network capacity – even if all road links are very long and all queue storage capacities are very large. This is done by giving an example:

a simple two-route signalised network is specified and it is shown that, with spatial queueing, equilibria consistent with P_0 do not exist for many feasible demands even if the queue storage capacities are very large.

This negative P_0 result shows the importance of considering the "spatial" vector $\mathbf{k}$ in $P_{\mathbf{hk}}$. The paper also shows that:

in the same simple two-route signalised network, with spatial queueing, equilibria consistent with $P_{\mathbf{hk}}$ do exist for any feasible demands provided queue storage capacities are sufficiently large, if $\mathbf{k}$ is correctly chosen.

An explanation of the capacity-maximising property of P_0

The capacity maximising property of P_0 is as follows. Suppose given a capacitated model which allows route choice. Suppose also that there is a given inelastic demand D which can be met by some (typically unknown) signal timings and some (typically unknown) route-flows; then under additional conditions there is a

[*route-flow pattern and signal timing*] pair for which

1. the routeing satisfies Wardrop's equilibrium condition and
2. the signal timings satisfy P_0.

This holds under various different "additional conditions" in various different models: (a) with a Beckmann-style cost function in a steady state flow model, (b) with "vertical" queueing in the quasi-dynamic model in Smith et al. (2019a), and (c) with *certain* spatial queueing models also in Smith et al. (2019a). Yet the policy requires no knowledge of the origin – destination distribution and so may be implemented locally and automatically.

Under certain, further, conditions the above equilibrium [route-flow pattern and signal timing] pair will be stable: natural route swapping and green-time swapping will converge to the above pair. Further research on this stability topic is underway.

These properties of P_0 make the policy ideal for the digital age and so important. There are many opportunities for development of this policy especially using pricing and signal control together (Smith et al., 2022).

A road map of topics considered in the following sections of the paper

The following sections and their central elements are as follows.

Section 2: **A fixed demand capacitated traffic assignment model with spatial queues.**

This section outlines the general assignment / control model utilised and also specifies:

the simple network studied in the paper,
three responsive control policies P_0, $P_{\mathbf{h}}$ and $P_{\mathbf{hk}}$,
two link performance models, one with vertical queues and one with spatial queues, and
five equilibrium / control performance tests of the control policies P_0, $P_{\mathbf{h}}$ and $P_{\mathbf{hk}}$.

Section 3. **The five equilibrium / control performance results.**
This gives the performances of the three control policies $\boldsymbol{P_0}$, $\boldsymbol{P_h}$ and $\boldsymbol{P_{hk}}$ in tests 1, 2, 3a, 3b and 4.
Section 4. Graphs of the performance results of tests 1, 2, 3a, 3b and 4.
This shows graphs comparing the performances of the three policies P_0, P_h and P_{hk} in tests 1, 2, 3a, 3b and 4.
Section 5. ***The Daganzo network with signal control and the three policies P_0, P_h and P_{hk}.***
Here this study is connected to previous research in Daganzo (1998), which gives a network where a lack of queueing space causes blocking back; this prevents equililibrium for certain feasible demands.
Section 6. **This study in the context of modern information systems and real-time control**
This section discusses the equilibrium modelling in this paper in the context of greater data availability.
Section 7. **Conclusion.**
This section gives conclusions and opportunities for further work.

A fixed demand capacitated traffic assignment model with spatial queues

The route-choice principle adopted throughout this paper follows the familiar Wardrop (1952) equilibrium notion: for each origin-destination pair, more costly routes are not used. This Wardrop equilibrium traffic assignment principle may also be expressed using link flows and bottleneck or queueing delays instead of route flows. In this paper we utilise a link formulation of "equilibrium with queueing delays"; this formulation is given in detail in Smith et al. (2019a) and Smith et al. (2022), and, in less detail, later in this paper.

One of the aims of this study is to analyse traffic control policies for planning applications. We do not consider signal phase sequences in detail, neither the detailed dynamics of the queuing process caused by the signal process (queue formation and dissipation during red and green phases, front and back of queue shockwave dynamics, etc.). Instead, the models described in this paper may be thought of as seeking to model the peak of a peak period where queues are at their maximum, without modelling the building up or the decay of those queues.

Moreover, our modelling approach is fully deterministic, hence we do not consider overflow queues at understaturated conditions caused by the stochastic nature of the arrival and departure processes and due to the probability of vehicles arriving during red or green phases (Viti and Van Zuylen, 2010), but we assume that queues and delays occur only at oversaturated signals. These are considered acceptable assumptions considering that the aim of the study is to understand how to make the best use of available network capacity. We do suggest further work to include more signal control detail in the conclusion of the paper.

Finally, in this paper we will, in the main, suppose that the queueing capacity of all links is very large but finite.

A basic link model with a spatial representation of queues and a finite queue storage capacity

As often in traffic modelling, each real-life traffic lane is here represented by:

a node which represents the entry point of the lane,
a node which represents the exit point or the stop line of the lane, and
a directed link joining these two nodes, which represents the stretch of lane between the entry and the exit of the lane.

A graphical representation of link i is shown in Fig. 1. The flow along link i is to be v_i vehicles per second; the saturation flow at the exit of link i is s_i vehicles per second; the queue originating at the exit of link i contains Q_i vehicles; the maximum possible value of Q_i is $MaxQ_i$ (fixed for a link i); and the freeflow travel time to traverse the entire length of link i (when the queue $Q_i = 0$ and the flow is v_i) is c_i. The link i "state" may be thought of as (v_i,Q_i) and this link state 2-vector is to be confined to a set of supply-feasible pairs as follows:

$$v_i \leq s_i \text{ and } Q_i \leq MaxQ_i.$$

Throughout the rest of the paper, each link cost c_i is to be a positive constant and link costs will be measured in seconds. All travellers have the same value of time.

Suppose that the saturation flow s_i at the link i exit is the only constraint on the link flow v_i (vehicles per second). To calculate the queueing or bottleneck delay b_i (seconds per vehicle) at the link i exit, it has often been proposed that (see for example Thompson and Payne, 1975; Smith et al. 2019b, 2022):

$$b_i = \frac{Q_i}{s_i}. \tag{1}$$

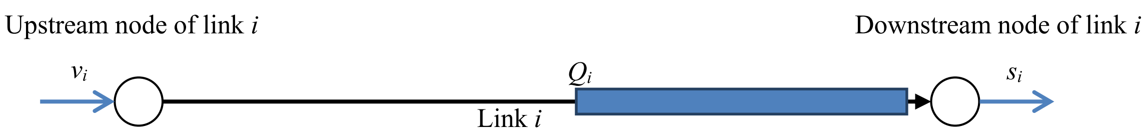

Fig. 1. A single link model representing a single traffic lane.

The whole time t_i of travel along link i has then often been written, by for example Thompson and Payne (1975); Smith et al. (2019a), as follows:

$$t_i = c_i + b_i. \tag{2}$$

This formula (2) may be regarded as the definition of "vertical queueing delay". Eq. (2) says that the total travel time taken to exit link i equals the time taken to traverse the whole length of link i added to the bottleneck delay at the link i exit. There are two interpretations of this formula (2). First: the bottleneck queue is here only entered when the whole link length has been traversed and the queue must then be thought of as being vertical. Second, Eq. (2) may also be thought of as a definition of the bottleneck delay b_i if the link traversal time $t_i = c_i + b_i$ can be measured (perhaps using Bluetooth) and c_i is known; see Mercader et al. (2020).

Related studies have been done by Lawson et al. (1997) who discuss two interpretations of "queueing delay" in dynamic traffic flow modelling: this discussion also applies in our case here where we have green-times g_i. In addition to the delay b_i above, which is the excess travel time felt by the vehicle, they discuss the time a vehicle is in a queue: with spatial queueing this queueing delay must be larger (perhaps much larger) than the excess travel time b_i above. Lawson et al. (1997) point out that social costs (such as pollution and the space taken up by queueing vehicles) depend greatly on the *spatial* queueing delay and depend much less on the excess travel time delay.

To estimate this *spatial queueing delay,* we consider three cases.

First suppose that $Q_i = 0$. In this case (using (1)), $b_i = 0.$, and formula (2) is entirely reasonable and the travel time simply consists of the traversing time of the whole link c_i.

Second, suppose that $Q_i = MaxQ_i$. The queue will, in this case, take up all the available space on link i and so if b_i now represents the horizontal queuing delay, formula (2) overestimates the travel time for a link; in this case the formula should be: $t_i = b_i$ and not $t_i = c_i + b_i$.

Third, suppose now that $0 < Q_i < MaxQ_i$. In this case, in real life, the queue on link i covers part of its length and only the remainder has to be traversed (with no queue). The queue will, in this case, take up some space on link i and so if b_i now represents the horizontal queuing delay, formula (2) overestimates the travel time for a link; the overestimation is greater the larger the real, spatially extensive, queue is and the error is maximum when the queue reaches the upstream node. At this point the formula should be: $t_i = b_i$ and not $t_i = c_i + b_i$.

To create a more realistic formula than the vertical queueing formula (2), so that b_i represents real-life spatial queueing, we suppose that the head of the queue on link i is always adjacent to the link exit and that the total travel time t_i spent in traversing link i is given by:

$$t_i(b_i) = a_i c_i + b_i; \tag{3}$$

a sum of non-queueing travel time $a_i c_i$ and a queueing travel time $b_i$seconds. Here, in (3), the "shrinkage factor" $a_i \leq 1$ is to be chosen so as to take some careful but simple account of the fact that as the horizontal queueing delay b_i (and hence the horizontal queue length or volume Q_i) grows, the unqueued link length (and hence the non-queueing travel time $a_i c_i$) shrinks.

In this paper we put:

$$a_i = \frac{MaxQ_i - Q_i}{MaxQ_i}. \tag{4}$$

If $Q_i = MaxQ_i$, then $a_i = 0$ and all the time spent traversing link i using (3) becomes entirely queueing time b_i. Also if $Q_i = 0$ then $b_i = 0$ and $a_i = 1$, and all the time spent traversing link i using (3) becomes entirely running time c_i. Both are reasonable, and intermediate cases where $0 < Q_i < MaxQ_i$ are also reasonable. In this intermediate case the $Q_i$vehicles in a queue at the link exit will occupy a proportion $\frac{Q_i}{MaxQ_i}$ of the length of link i, leaving only the proportion $1 - \frac{Q_i}{MaxQ_i}$ of the length of link i to be traversed without queueing. This will take $\left(1 - \frac{Q_i}{MaxQ_i}\right)c_i$ seconds to traverse since the traversal time of the whole length of link i is c_i seconds.

Substituting $a_i = \frac{MaxQ_i - Q_i}{MaxQ_i} = 1 - \frac{Q_i}{MaxQ_i}$ into (3), we obtain:

$$t_i(b_i) = \left(1 - \frac{Q_i}{MaxQ_i}\right)c_i + b_i = c_i - \frac{Q_i}{MaxQ_i}c_i + b_i = c_i - \frac{b_i s_i}{MaxQ_i}c_i + b_i = c_i + \left(1 - \frac{s_i c_i}{MaxQ_i}\right)b_i.$$

Thus, this value of a_i in (3) leads to the following equations:

$$t_i(b_i) = c_i + k_i b_i \tag{5a}$$

where

$$k_i = 1 - \frac{s_i c_i}{MaxQ_i}. \tag{5b}$$

We now may naturally generalise (5a, b) to take account of a green-time proportion g_i at the link exit (as well as the exit saturation flow s_i). The corresponding formulae are:

$$t_i(g_i, b_i) = c_i + k_i(g_i)b_i \tag{5c}$$

where k_i now depends on g_i as follows:

$$k_i = k_i(g_i) = 1 - \frac{s_i g_i c_i}{MaxQ_i}. \quad (5d)$$

Formula (5d) defines a shrinkage factor, which has a central role in our study. Since its definition contains both saturation flow and queue capacities (i.e. buffer spaces), it allows us to explicitly consider both physical constraints characterising horizontal queueing processes.

A new road parameter, θ_i, and a link feature captured by θ_i

Motivated by (5c,d), for each link *i* we let

$$\theta_i = \frac{s_i c_i}{MaxQ_i} = \frac{c_i}{MaxQ_i/s_i} = (\text{link } i \text{ freeflow travel time}) \Big/ (\text{maximum link } i \text{ queueing time when } g_i = 1). \quad (6)$$

We now suppose, partly for simplicity in this paper, that $0 < \theta_i \leq 1$ for all links *i*. This implies (for all links *i*) that the link *i* freeflow travel time is never more than (the maximum link *i* queueing time when $g_i = 1$). This is a very reasonable condition since it is unlikely that vehicles would move more quickly while moving in a queue than in free flow conditions. (However $\theta_i > 1$ may arise if vehicles form a "train" or a tightly controlled platoon. We do not consider this possibility in this paper; we assume throughout this paper that $0 < \theta_i \leq 1$ for all links *i*.)

The parameter θ_i, by definition, relates the minimum time a vehicle needs to traverse the link in free flow conditions, with the time needed by the same vehicle to traverse the link when the queue has completely filled the storage capacity. Hence, the parameter includes the main infrastructural design parameters for link *i*, namely the link staturation flow s_i, the link storage capacity $MaxQ_i$ and the link free flow travel time c_i.

Here, in Fig. 2, we give three example links (*1, 2, 3*) to illustrate how the value of θ_i reflects or represents the shape of link *i*. Each of the three links here in Fig. 2 has the same length, the same freeflow speed or freeflow travel time, the same link-exit-width and the same link-exit saturation flow.

But the three road links, *1, 2, 3*, in Fig. 2 differ in the widths (and hence the areas and queue storage capacities) of the link upstream of the link exits. The average width of link *1* is less than the average width of link *2* which is less than the average width of link *3*. Thus link *1* has a smaller area and maximum queue, $MaxQ_i$, than link *2* which has a smaller area and maximum queue than link *3*. These three links may represent a part of a real-life road section where the number of lanes increases downstream (e.g. from one to two lanes), a section where the number of lanes remains constant, and a section where there is one or more lanes dropped (e.g. from three to two lanes), respectively.

So, since the links have the same freeflow travel time c_i and the same exit saturation flow s_i, the value of θ_i (given by (6)) for link *1* in Fig. 2 will be larger than the value of θ_i for link *2* which will in turn be larger than the value of θ_i for link *3*.

A simple example

It is interesting to compare five [control policy / queueing model] combinations in a simple example; to do this we utilise the Fig. 3 network. Let

s_i = the saturation flow at the link *i* exit (in vehicles per second: $i = 1, 2, 3$);
c_i = the free flow cost/time of travel via link *i* (in seconds; constant: $i = 1, 2, 3$);
C_i = the free flow cost/time of travel via route *i* (in seconds; constant: $i = 1, 2$);
x_i = the flow along link *i* (in vehicles per second: $i = 1, 2, 3$);
b_i = the bottleneck delay felt at the link *i* exit (in seconds: $i = 1, 2$);
T = the steady total inflow at the origin, heading for the destination (v/s);
$MaxQ_i$ = the maximum queue possible on link $i = 1, 2$; and, following (6),
$\theta_i = c_i/(MaxQ_i/s_i) = c_i s_i/MaxQ_i$ for $i = 1, 2$.

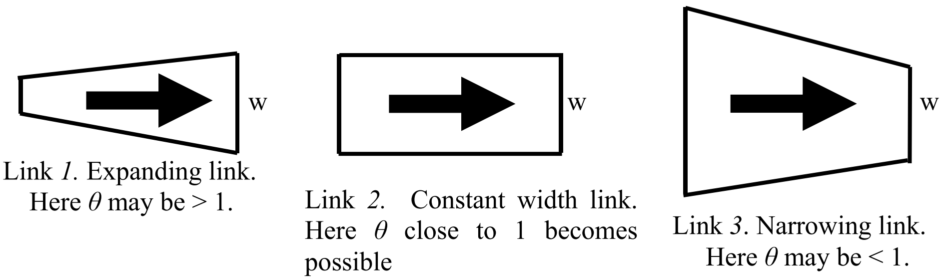

Fig. 2. Three road links with the same lengths, outflow widths w (metres) and saturation flows; but different shapes. The expanding road link *1* has a smaller area and so a smaller maximum queue, a smaller maximum queueing time and so a larger value of θ than the rectangular link *2*. The rectangular road link *2* has a smaller area and so a smaller maximum queue, a smaller maximum queueing time and so a larger value of θ than the narrowing link *3*.

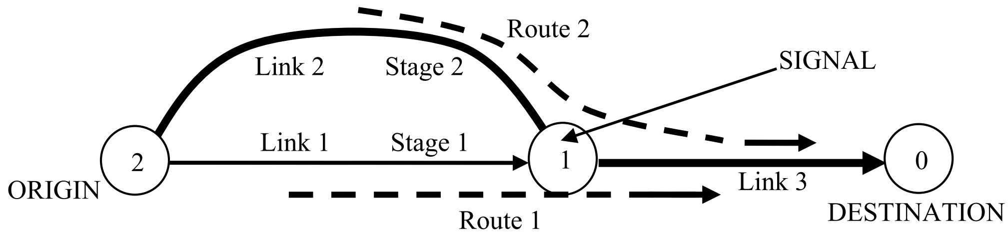

Fig. 3. A simple two route signal-controlled network; link 2 is wider and longer than link 1.

We suppose that c_1, c_2 and c_3 are constant, that $C_1 = c_1 + c_3$ and $C_2 = c_2 + c_3$.
Throughout we suppose that

$$0 < s_1 < s_2 < s_3,$$

$$0 < \theta_1 = \theta_2 = \theta \leq 1,$$

and$c_2 = c_1 + \Delta$where $\Delta > 0$. We also define

$$r = s_1/s_2$$

so that $r < 1$. Here, with the above definition (6) of θ_i and our assumption that $\theta_1 = \theta_2 = \theta$, (5d) becomes:

$$k_i = k_i(g_i) = 1 - \frac{s_i g_i c_i}{MaxQ_i} = 1 - \frac{s_i c_i}{MaxQ_i} g_i = 1 - \theta g_i.$$

Definition of spatial queueing equilibrium consistent with P_0 on this network

Given a steady demand T (v/s) satisfying $0 < T < s_2$; we now consider (**x, g, b**) where

$$\mathbf{x} = (x_1, x_2, x_3),\ \mathbf{g} = (g_1, g_2) \text{ and } \mathbf{b} = (b_1, b_2).$$

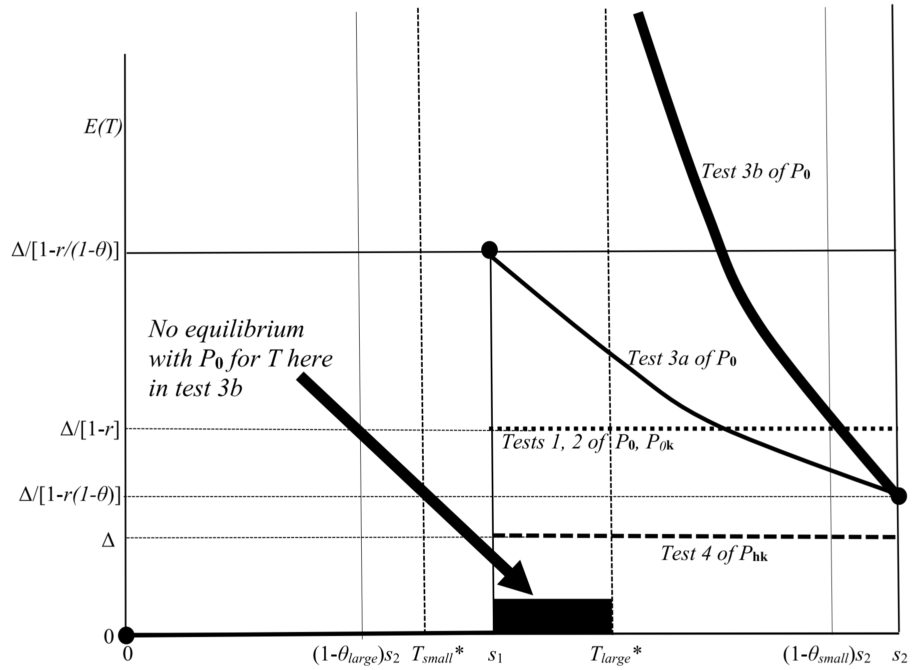

Fig. 4. The graphs show how $E(T)$ varies with T in tests 1, 2, 3a, 3b, and 4 when T increases from 0 to s_2. For all tests $E = 0$ when $T < s_1$. In tests 1, 2, 3a and 4, E jumps upwards from 0 at $= s_1$; the jump is greatest when the policy is P_0 and is least with $P_{\mathbf{hk}}$, the biased spatial version of P_0. The test 3a performance results are given by the lower solid curve. The test 3b performance results are given by the thicker upper solid curve; *there is a gap here because there is no consistent equilibrium when $s_1 < T < T^*_{large}$. This range of T is indicated by the thick short black line on the horizontal or T axis.*

Definition. (**x, g, b**) is a *spatial queueing equilibrium consistent with* P_0 if and only if

1 **x** is demand feasible or satisfies the given demand or $x_1 + x_2 = T = x_3$, $x_1 \geq 0$, $x_2 \geq 0$;
2 (**x, g**) is supply-feasible or $x_1 \leq s_1 g_1$ and $x_2 \leq s_2 g_2$ and $g_1 + g_2 = 1$;
3 (**x, b**) is a spatial Wardrop equilibrium so that if $s_1 < T < s_2$ then $c_1 + k_1 b_1 = c_2 + k_2 b_2$;
4 (**x, g, b**) satisfies P_0 so that if $s_1 < T < s_2$: $s_1 b_1 = s_2 b_2$ and
5 (**x, g, b**) is a queueing equilibrium or ($b_i = 0$ if $x_i < s_i g_i$).

Notes:

1 The last of these conditions, condition (5), says that the bottleneck delay and the queue length stay at zero if the link is unsaturated. This condition is more severe than the usual assumption for signal control. This condition is used here as it enables a relatively simple analysis. We believe that this condition may be relaxed, while obtaining similar results, by letting s_i depend on b_i. But we leave that to further work. In this particular case here this assumption is used only to derive the flat part of the $E(T)$ graphs in Fig. 4 where $T < s_1$.
2 In (3) above, if $k_i = 1$ the queueing is said to be vertical.
3 In (3) above, k_i depends on g_i.
4 The definition of P_0 in (4) above is given in more detail in Smith et al. (2019a, b, 2022).

The five equilibrium / control performance tests

The "extra (or excess) travel time $E = E(T)$" will be utilised as the performance measure in each test, where

$$E = E(T) = (\text{the travel time at equilibrium minus } C_1)(s).$$

This will be the only performance measure for all five tests. In each test E will depend importantly on the constant origin-destination flow T (v/s), on $\Delta = c_2 - c_1 > 0$, on the control policy employed, and also, very importantly, on the queueing model (vertical or spatial) employed.

Throughout the performance tests 1, 2, 3a, 3b and 4 we assume that

$$0 < \theta \leq 1 \text{ and } k_i = k_i(g_i) = 1 - \theta g_i \text{ (for } i = 1, 2).$$

In test 4 we also assume that $\mathbf{h} = (s_1 c_1, s_1 c_2)$.

The 5 tests are as follows (Here $r = s_1/s_2 < 1$):

Test 1. Equilibrium performance of the P_0 *control policy (choose* **g** *so that* $s_1 b_1 = s_2 b_2$*) with vertical queueing;*

Test 2. Equilibrium performance of spatial P_0, $P_{0\mathbf{k}}$ *(choose* **g** *so that* $k_1 s_1 b_1 = k_2 s_2 b_2$*) with spatial queueing;*

Test 3a. Equilibrium performance of P_0 *(choose* **g** *so that* $s_1 b_1 = s_2 b_2$*) with spatial queueing when* $\theta = \theta_{small}$ *where* $\theta_{small} < 1 - r$*;*

Test 3b. Equilibrium performance of P_0 *(choose* **g** *so that* $s_1 b_1 = s_2 b_2$*) with spatial queueing when* $\theta = \theta_{large}$ *where* $\theta_{large} > 1 - r$*;*

Test 4. Equilibrium performance of $P_{\mathbf{hk}}$ *(choose* **g** *so that* $k_1 s_1 b_1 + h_1 = k_2 s_2 b_2 + h_2$*), biased spatial* P_0 (see Smith et al. (2019b) for its derivation) *with spatial queueing.*

Notes:

1. θ_{small} and θ_{large} are two possible values of θ in relation to the ratio s_1/s_2 of the saturation flows of the two links merging at the signalized intersection; the test results in these two tests *3a* and *3b* are very different.
2. Policies $P_{\mathbf{h}}$ and $P_{\mathbf{hk}}$ are variations on P_0 introduced in Smith et al. (2019b). $P_{\mathbf{h}}$ has improved performance relative to P_0 and $P_{\mathbf{hk}}$ has improved performance relative to P_0, and also takes account of spatial rather than vertical queueing.

The following two link queueing models (*LQMs*) will be considered in these 5 tests:

LQM1. Vertical queueing: Here we suppose that the total time $t_i(b_i)$ of traversing link i is $c_i + b_i$.

LQM2. Spatial queueing: Here we suppose that the total time $t_i(b_i)$ of traversing link i is $c_i + k_i b_i$ where $k_i = 1 - \theta g_i$. (This is an application to this example of (5a) and (5b), including green-time proportions.)

We now give, in Fig. 4 in Section 3, the equilibrium performance results for all of the five [control, queueing model] tests *1, 2, 3a, 3b, 4*. In the derivation of all these results we only consider $s_1 < T < s_2$ in detail.

See Smith et al. (2019b) for detailed justifications of the test results *1, 2* and *4*. These results are reproduced here in Fig. 4 for comparison with the results of tests *3a* and *3b*, which are the main concerns of this paper.

The five equilibrium / control performance results

In calculating the results of tests 1, 2, 3a, 3b, 4 below we only consider values of T satisfying $s_1 < T < s_2$.

Here $r = s_1/s_2 < 1$. Also we suppose in this example that $MaxQ_1$ and $MaxQ_2$ satisfy:

$$\frac{s_1 C_1}{MaxQ_1} = \frac{s_2 C_2}{MaxQ_2} = \theta$$

and that s_3 is very large, so there is no queueing on link *3*. Given *T* satisfying $s_1 < T < s_2$, at equilibrium both links 1 and 2 must be used and here we suppose that at equilibrium there must be (stationary) queues on both links 1 and 2. In this case, where *T* satisfies $s_1 < T < s_2$, the equilibrium green-time vector

$$\mathbf{g} = \left(\frac{s_2 - T}{s_2 - s_1}, \frac{T - s_1}{s_2 - s_1}\right).$$

Any other green-time vector does not support stationary queues on links 1 and 2 at equilibrium if $s_1 < T < s_2$. (See for example Smith et al. (2019b, 2022).)

1. Equilibrium performance of the P_0 control policy, with vertical queueing. In this case:

$$E(T) = (\text{travel time at equilibrium minus } C_1) = C_2 + b_2 - C_1 = C_1 + b_1 - C_1 = b_1 = b_2 + \Delta = \Delta/(1-r)$$

2. Equilibrium performance of spatial P_0, $P_{0\mathbf{k}}$ with spatial queueing. In this case:

$$E(T) = (\text{travel time at equilibrium minus } C_1) = C_2 + k_2 b_2 - C_1 = k_1 b_1 = k_2 b_2 + \Delta = \Delta/(1-r)$$

3. Equilibrium performance of P_0 with spatial queueing.

In this case, for any *T* which has an equilibrium consistent with spatial queueing and policy P_0:

$$E(T) = (\text{travel time at equilibrium minus } C_1) = \frac{\Delta}{\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2} - 1} + \Delta = \frac{\Delta \frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2}}{\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2} - 1}. \tag{7}$$

In (7), $\mathbf{g} = \left(\frac{s_2 - T}{s_2 - s_1}, \frac{T - s_1}{s_2 - s_1}\right)$. Other green-time vectors do not support stationary queues on links 1, 2.

Proof of this formula (7).

Suppose that *T* satisfies $s_1 < T < s_2$ *and that T has an equilibrium consistent with spatial queueing and policy P_0*. Then at any such equilibrium we have the following statements.

At a Wardrop equilibrium with spatial queueing: $C_1 + k_1 b_1 = C_2 + k_2 b_2$.

Thus at Wardrop equilibrium with spatial queueing: $k_1 b_1 = k_2 b_2 + \Delta$. (Here $\Delta = C_2 - C_1$.)

The P_0 policy ensures that: $s_1 b_1 = s_2 b_2$.

Therefore: $s_1 k_2 b_2 + s_1 \Delta = s_1 k_1 b_1 = k_1 s_1 b_1 = k_1 s_2 b_2$, and so: $s_1 \Delta = (k_1 s_2 - s_1 k_2) b_2$ or: $\Delta = ((k_1/k_2)(s_2/s_1) - 1) k_2 b_2$.

Thus: $k_2 b_2 = \Delta/((k_1/k_2)(s_2/s_1) - 1)$.

Now $k_i = 1 - \theta g_i$ for $i = 1, 2$. ($\mathbf{k}$ satisfies (5d).) So *for any T which satisfies $s_1 < T < s_2$ and has an equilibrium consistent with spatial queueing and policy P_0*:

$$\begin{aligned}
E(T) &= [\text{the travel time at equilibrium minus } C_1] \\
&= C_2 + k_2 b_2 - C_1 = k_2 b_2 + \Delta = \Delta/((k_1/k_2)(s_2/s_1) - 1) + \Delta \\
&= \frac{\Delta}{\frac{1}{r}\cdot\frac{1 - s_1 g_1 C_1/MaxQ_1}{1 - s_2 g_2 C_2/MaxQ_2} - 1} + \Delta \\
&= \frac{\Delta}{\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2} - 1} + \Delta = \frac{\Delta \frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2}}{\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2} - 1}.
\end{aligned} \tag{8}$$

QED.

We now consider two cases: *3a* and *3b*.

In *3a*, $0 < \theta = \theta_{small} < 1 - r$ and in *3b*, $1 - r < \theta = \theta_{large} \leq 1$.

In both of these two cases we utilise Theorem 1 below which is proved in the appendix.

Theorem 1. In the equilibrium model above, where $0 < \theta \leq 1$, suppose that T satisfies $s_1 < T < s_2$. Then there is an equilibrium consistent with spatial queueing and P_0 (and also the extra travel time E(T)is given by (7), (8)) if and only if

$$T > T^*(\theta) = \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/\theta}{s_1 + s_2} \tag{9}$$

3a. Existence of equilibrium and the equilibrium performance of P_0 ($s_1 b_1 = s_2 b_2$) with spatial queueing when

$$\theta = \theta_{small} < 1 - r \text{ or equivalently } (1-\theta)s_2 > s_1.$$

Suppose $\theta = \theta_{small} < 1 - r$ or $(1 - \theta_{small})s_2 > s_1$. Let the corresponding value of T^* (given in (9)) be T^*_{small}. It follows that in this case *3a*, where $\theta_{small} < 1 - r$ or $(1 - \theta_{small})s_2 > s_1$,

$$\begin{aligned} T^*_{small} &= \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/\theta_{small}}{s_1 + s_2} \\ &< \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/(1-r)}{s_1 + s_2} \\ &= \frac{s_1^2 + s_2^2 - s_2(s_2 - s_1)^2/(s_2 - s_1)}{s_1 + s_2} \\ &= \frac{s_1^2 + s_2^2 - s_2(s_2 - s_1)}{s_1 + s_2} \\ &= \frac{s_1^2 + s_1 s_2}{s_1 + s_2} = s_1 \end{aligned}$$

Thus $s_1 > T^*_{small}$, and since $s_1 < T < s_2$, it follows that $T > s_1 > T^*_{small}$ and so by Theorem 1 (proved in the appendix) there is an equilibrium consistent with spatial queueing and P_0. Further, $E(T)$ is given by (7) and (8).

3b. Existence of equilibrium and the equilibrium performance of P_0 ($s_1 b_1 = s_2 b_2$) with spatial queueing when $\theta = \theta_{large}$ where $\theta_{large} > 1 - r$ or equivalently $(1 - \theta_{large})s_2 < s_1$.

Suppose $\theta = \theta_{large}$ where $\theta_{large} > 1 - r$ or $(1 - \theta_{large})s_2 < s_1$. Let the corresponding value of T^* (given in Eq. (9)) be T^*_{large}.

It follows in this case *3b*, where $1 \geq \theta_{large} > 1 - r$ or $0 \leq (1 - \theta_{large})s_2 < s_1$, that

$$\begin{aligned} T^*_{large} &= \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/\theta_{large}}{s_1 + s_2} \\ &> \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/(1-r)}{s_1 + s_2} \\ &= \frac{s_1^2 + s_2^2 - s_2(s_2 - s_1)^2/(s_2 - s_1)}{s_1 + s_2} \\ &= \frac{s_1^2 + s_2^2 - s_2(s_2 - s_1)}{s_1 + s_2} \\ &= \frac{s_1^2 + s_1 s_2}{s_1 + s_2} = s_1 \end{aligned}$$

It also follows in this case that

$$\begin{aligned} T^*_{large} &= \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/\theta_{large}}{s_1 + s_2} \\ &\leq \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/1}{s_1 + s_2} \\ &= \frac{2 s_1 s_2}{s_1 + s_2} = s_2 \frac{s_1 + s_1}{s_1 + s_2} < s_2. \end{aligned}$$

Thus, in this *3b* case, where $1 \geq \theta_{large} > 1 - r$ or $0 \leq (1 - \theta_{large})s_2 < s_1$,

$$s_1 < T^*_{large} = \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/\theta_{large}}{s_1 + s_2} < s_2$$

and we now consider the two cases $s_1 < T < T^*_{large}$ and $T^*_{large} < T < s_2$.

Theorem 1 immediately yields the following results:

(i) if $s_1 < T < T^*_{large}$ then there can be no equilibrium consistent with P_0 and
(ii) if $T^*_{large} < T < s_2$ then there exists an equilibrium consistent with P_0.

The equilibrium in (ii) has performance $E(T)$ which is given by Eq. (7).

Spatial equilibrium performance of biased spatial P_0, $P_{\mathbf{hk}}$, where$\mathbf{h} = [s_1 c_1, s_1 c_2]$, $k_i = 1 - \theta g_i$.

In this case:

$$E = (\text{the travel time at equilibrium minus } C_1) = C_2 + k_2 b_2 - C_1 = k_2 b_2 + \Delta = (\Delta - (h_2 - h_1)/s_1)/((s_2/s_1) - 1) + \Delta = \Delta.$$

Graphs of the performance results of tests 1, 2, 3a, 3b and 4

Comparisons of the equilibrium performance results 1, 2, 3a, 3b and 4

Now we compare the performance measure $E = E(T)$ in all five tests, as T increases from 0 to s_2. In this Fig. 4, $(1 - \theta)s_2 > s_1$ (in test *3a*) and $(1 - \theta)s_2 < s_1$ (in test *3b*) where $0 < \theta \le 1$.

An important feature of this whole picture in Fig. 4 is that we consider two values of θ: these are θ_{small} and θ_{large} where

$$\theta_{small} < 1 - r < \theta_{large} \text{ or equivalently } \left(1 - \theta_{large}\right)s_2 < s_1 < (1 - \theta_{small})s_2.$$

Focussing first on test *3a*, when $\theta_{small} < 1 - r$ or equivalently $s_1 < (1 - \theta_{small})s_2$, the lower curved bold line in Fig. 4 shows that if responsive P_0 policy is utilised when spatial queueing occurs and $\theta_{small} < 1 - r$ or equivalently $s_1 < (1 - \theta_{small})s_2$, then the extra delay E will, at equilibrium when T is only slightly larger than s_1, be $\Delta/(1 - r/(1 - \theta_{small}))$; which may be much greater than $\Delta/(1 - r)$ seconds (which is the value of E with vertical queueing when T is just slightly larger than s_1; see Test 1).

The lower curved bold line in Fig. 4 shows also that, when $T > s_1$, $E(T)$ *decreases* with T. This decreasing behaviour of $E(T)$may seem counterintuitive; it is the result of the three-way interaction between (i) the responsive P_0 policy, (ii) routeing decisions by drivers, and (iii) the spatial queueing model adopted.

The horizontal dotted lines in Fig. 4 also show how the spatial queueing performance is improved (compared to the P_0 performances in both tests *3a* and *3b,* shown by the solid curved lines) by switching to $P_{0\mathbf{k}}$, the spatial version of P_0, and how performance is further improved by switching to the biased spatial version of P_0, namely $P_{\mathbf{hk}}$, with suitable **h** and **k**.

Focussing now on test *3b*, when $\theta_{large} > 1 - r$ or equivalently $s_1 > (1 - \theta_{large})s_2$, the upper bold line in Fig. 4 shows that if responsive P_0 policy is utilised when spatial queueing occurs and $\theta_{large} > 1 - r$ or equivalently $s_1 > (1 - \theta_{large})s_2$, then $T^*(\theta_{large}) > s_1$. Now the extra delay $E(T)$ will become increasingly large as T decreases toward $T^*(\theta_{large})$ (which is $> s_1$). *There is no consistent equilibrium when* $s_1 < T \le T^*(\theta_{large})$.

The Daganzo network with signal control and the three policies P_0, $P_{0\mathbf{k}}$ and $P_{\mathbf{hk}}$

The study we have presented in this paper has analogies with previous research by Daganzo (1998), which gives a network where a lack of queueing space causes blocking back; this prevents equilibrium in a spatial queueing model for certain feasible demands.

Here, following Daganzo, we now assune that $MaxQ_i$ is *not large* in our Fig. 1 network. We show how the three policies P_0, $P_{0\mathbf{k}}$ and $P_{\mathbf{hk}}$ perform on the network in Fig. 1 in this case.

QUESTIONS: When T is slightly greater than s_1, what are the equilibrium link 1 delay and the equilibrium link 1 queue for these three policies?

Table 1. The top row shows the three traffic control policies P_0 (with two cases), $P_{0\mathbf{k}}$ and $P_{\mathbf{hk}}$. The figures in column 3, 4 and 5 give the equilibrium delay and the equilibrium queue for these three policies when T is just larger than s_1. With P_0 there are two cases: $\theta < 1 - r$ and $\theta > 1 - r$. If $\theta > 1 - r$ and T is slightly larger than s_1 there is no equilibrium delay consistent with P_0, and also no equilibrium queue.

Now

$$\Delta/(1 - r/(1 - \theta))\rangle\Delta/(1 - r) > \Delta.$$

Using these two inequalities and the results in Table 1 we obtain, for T slightly larger than s_1,

(A) If $\theta > 1 - r$ there is no equilibrium consistent with P_0;
(B) If $\theta < 1 - r$, P_0 has a larger equilibrium delay on link 1 and requires more queueing space than $P_{0\mathbf{k}}$, and
(C) $P_{0\mathbf{k}}$ has a larger equilibrium delay and requires more queueing space than $P_{\mathbf{hk}}$.

This study in the context of modern information systems and real-time control

Equilibrium modelling with signal control has typically involved equilibrium route choice modelling for fixed signals, combined with rather slow signal control adjustments to change those fixed signal settings. For example, see Allsop (1974); Smith and van Vuren (1993) discuss day-to-day iterations involving signal control updates on one day followed by routeing adjustments on the next day. So it is reasonable to ask whether our equilibrium discussions in this paper are relevant when:

Table 1
shows the equilibrium delay and the equilibrium queue for these three policies when T is slightly greater than s_1.

	P_0 ($\theta > 1 - r$)	P_0 ($\theta < 1 - r$)	$P_{0\mathrm{k}}$	P_{hk}
Link *1* equilibrium delay b_1 when T is slightly larger than s_1 (see the vertical axis in Fig. 4)	There is no link 1 equilibrium delay	$\Delta/[1-r/(1-\theta)]$	$\Delta/[1-r]$	Δ
Link *1* equilibrium queue Q_1 when T is slightly larger than s_1.	There is no link 1 equilibrium queue	$s_1\Delta/[1-r/(1-\theta)]$	$s_1\Delta/[1-r]$	$s_1\Delta$

(A) information is now much more up-to-date and available, and
(B) signal responses are now much quicker and potentially much more discriminating.

Because of (A) and (B) above, modern signal control systems have a much greater potential impact than historic fixed time systems, and yet often modern traffic control systems have been, and are, used with policies which do not take proper account of route choice. This deficiency may now be having a much greater impact than previously when signals were more often fixed time.

So perhaps it is now more important to understand how to choose signal settings which take good account of route choice. With such understanding, currently powerful control systems would perhaps be much more productive in controlling traffic conditions while allowing properly for route (and mode) choices.

Local responsive control policies P_0, P_{0k} and P_{hk} have all been designed to take some proper account of route choice. Our study here shows that P_{0k} or P_{hk} are may well be much better than P_0. Fig. 4 shows that the equilibrium performnce disparities are very large even on this very simple network.

Conclusion

The central model discussed in this paper is a steady-state, link-based, fixed (or inelastic) demand equilibrium model with explicit link-exit capacities, explicit spatial-queueing delays, explicit (but large) bounds on queue storage capacities, and green-times. The model is a spatial quasi-dynamic assignment/control model. The main link model at the heart of this equilibrium model takes some account of the space taken up by queues. We have called this link model a "spatial queueing" model.

In the paper we have considered a simple example network and have given the results in Smith et al. (2019a) involving the three responsive control policies: P_0, P_{0k} (spatial P_0) and P_{hk} (biased spatial P_0) using a vertical queueing model and a spatial queueing model. We have also given the results of one new test, test *3b*, and we have illustrated the results of all five tests 1, 2, 3a, 3b and 4 in Fig. 4.

Test *3a* in this paper involves spatial queueing and the P_0 result obtained in test 3a depends on the inequality

$$\theta < 1 - r \text{ or } (1-\theta)s_2 > s_1.$$

We have shown that, in this case, each feasible value of the demand T gives rise to a (unique) extra travel time $E(T)$. So P_0 maximises the capacity of this network if $\theta < 1 - r$. We have in this paper also shown, in test *3b*, that if, on the other hand,

$$\theta > 1 - r \text{ or } (1-\theta)s_2 < s_1,$$

then $T^*(\theta) > s_1$ and there is a non-empty range R of the fixed demand T satisfying:

if T belongs to R then T is feasible but there is no equilibrium consistent with policy P_0.

$R = (s_1, T^*(\theta))$, the set of those T between s_1 and $T^*(\theta)$ in Fig. 4. So the control policy P_0 *does not maximise the capacity of this simple network in this case.*

Thus we have shown that P_0 does not always maximise the capacity of this simple network. There are many opportunities for further work: for example we might consider networks with links having differing values of θ_i and we might consider representing different link characteristics such as surface quality.

The main implication of this work is that when there is spatial queueing, as there must be in reality, it is important for this simple network to use either P_{0k} (spatial P_0) or P_{hk} (biased spatial P_0) rather than P_0. We believe that this result is likely to hold also for many other networks.

There are very many opportunities for further work; for example:

(1) more general networks need to be considered,
(2) stochasticity and overflow queues need consideration,
(3) inflow control to suit P_{0k} or P_{hk} merit consideration to extend this study to large peak-hour demands, and
(4) the stability of these policies has never been studied and this represents another opportunity.

CRediT authorship contribution statement

Michael J Smith: Conceptualization, Methodology, Validation, Writing – original draft, Writing – review & editing. **Francesco Viti:** Conceptualization, Methodology, Writing – review & editing. **Wei Huang:** Conceptualization, Methodology, Writing – review & editing. **Richard Mounce:** Conceptualization, Methodology, Writing – review & editing.

Appendix

Theorem 1. In the model above, where $0 < \theta \leq 1$, *suppose that T satisfies* $s_1 < T < s_2$. *Then there is an equilibrium consistent with spatial queueing and* P_0 *and also the extra travel time* $E(T)$ *is given by (7) if and only if*

$$T > T^*(\theta) = \frac{s_1^2 + s_2^2 - (s_2 - s_1)^2/\theta}{s_1 + s_2}.$$

Proof.

Fix T where $s_1 < T < s_2$. Put $\mathbf{g} = ((s_2 - T)/(s_2 - s_1),(T - s_1)/(s_2 - s_1))$and

$$F(T) = \frac{1-\theta g_1}{1-\theta g_2} = \frac{1-\theta(s_2-T)/(s_2-s_1)}{1-\theta(T-s_1)/(s_2-s_1)}.$$

Now $s_1 < T < s_2$ and $0 < \theta \le 1$, so

$$1-\theta(s_2-T)/(s_2-s_1)\rangle 0 \text{ and } 1-\theta(T-s_1)/(s_2-s_1)\rangle 0.$$

It follows that $F(T)$ is well-defined and > 0 and hence that

$\frac{\Delta \frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2}}{\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2}-1}$ is well-defined and positive if and only if $\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2} - 1 > 0$.

Under what conditions is $\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2} - 1 > 0$?

$\frac{1}{r}\cdot\frac{1-\theta g_1}{1-\theta g_2} - 1 > 0$ if and only if $F(T) = \frac{1-\theta(s_2-T)/(s_2-s_1)}{1-\theta(T-s_1)/(s_2-s_1)} > r = \frac{s_1}{s_2}$, which happens if and only if $s_2(1-\theta(s_2-T)/(s_2-s_1)) > s_1(1-\theta(T-s_1)/(s_2-s_1))$(since $1-\theta(T-s_1)/(s_2-s_1) > 0$), which happens if and only if $(s_1+s_2)\theta T/(s_2-s_1) > s_1(1+\theta s_1/(s_2-s_1)) - s_2(1-\theta s_2/(s_2-s_1))$, which happens if and only if $(s_1+s_2)\theta T/(s_2-s_1) > (s_1-s_2) + \theta(s_1^2+s_2^2)/(s_2-s_1)$, which happens if and only if $T > (s_1^2 + s_2^2 - (s_2-s_1)^2/\theta)/(s_1+s_2)$.

Let

$$T^* = T^*(\theta) = \left(s_1^2 + s_2^2 - (s_2-s_1)^2/\theta\right)/(s_1+s_2).$$

It follows then that there is an equilibrium consistent with spatial queueing and policy P_0 with performance given by (7) and (8) if and only if $T > T^*(\theta)$. QED

References

Allsop, R.E., 1971. Delay-minimizing settings for fixed-time traffic signals at a single road junction. IMA J. Appl. Math. 8 (2), 164–185.

Allsop, R.E., 1974. Some possibilities for using traffic control to influence trip distribution and route choice. In: Buckley, D.J. (Ed.), Proceedings of the Sixth International Symposium on Transportation and Traffic Theory. Elsevier, Sydney, pp. 345–374.

Bliemer, M.C.J., Raadsen, M.P.H., Smits, E.S., Zhou, B., Bell, M.G.H., 2014. Quasi-dynamic traffic assignment with residual point queues incorporating a first order node model. Transp. Res. Part B 68, 363–384.

Bliemer, M.C.J., Raadsen, M.P.H., 2017. Static traffic assignment with residual queues and spillback. In: Proceedings of 17thSwiss Transportation Research Conference. Transport and Mobility Laboratory, EPF Lausanne, Monte Verita, Switzerland, pp. 1–32.

Daganzo, C.F., 1998. Queue spillovers in transportation networks with a route choice. Transp. Sci. 32 (1), 3–11.

Gartner, N., 1977. Influence of control measures on traffic equilibrium in an urban highway network. Transp. Res. Rec. 644, 125–129.

Gentile, G., Meschini, L., Papola, N., 2007. Spillback congestion in dynamic traffic assignment: a macroscopic flow model with time-varying bottlenecks. Transp. Res. Part B 41 (10), 1121–1138.

Han, K., Gayah, V., Piccoli, B., Friesz, T.L., Yao, T., 2014. On the continuum approximation of the on-and-off signal control on dynamic traffic networks. Transp. Res. Part B 61, 73–97.

Han, K., Gayah, V.V., 2015. Continuum signalised signal model for dynamic traffic networks. Transp. Res. Part B 77, 213–239.

Haddad, J., Zheng, Z., 2018. Adaptive perimeter control for multi-region accumulation-based models with state delays. Transp. Res. Part B. https://doi.org/10.1016/j.trb.2018.05.019.

Lam, W.H.K., Zhang, Y., 2000. Capacity-constrained traffic assignment in networks with residual queues. J. Transp. Eng. 126 (2), 121–128.

Lawson, T.W., Lovell, D.J., Daganzo, C.F., 1997. Using input-ouput analysis to determine spatial and temporal extents of a queue upstream of a bottleneck. Transp. Res. Rec. 1572, 140–147.

Levin, M.W., Hu, J., Odell, M., 2020. Max-pressure signal control with cyclical phase structure. Transp. Res. Part C 120, 102828.

Mercader, P., Uwayid, W., Haddad, J., 2020. Max-pressure traffic controller based on travel times: an experimental analysis. Transp. Res. Part C 110, 275–290.

Smith, M.J., 1979a. A local traffic control policy which automatically maximises the overall travel capacity of an urban road network. In: Proceedings of the International Symposium on Traffic Control Systems. University of California, Berkeley, pp. 298–302. Volume 2A, 11-32; and in *Traffic Engineering and Control*, 21, 1980.

Smith, M.J., 1979b. The existence, uniqueness and stability of traffic equilibria. Transp. Res. Part B 13 (4), 295–304.

Smith, M.J., 1987. Traffic control and traffic assignment in a signal controlled network with queueing. In: Gartner, N.H., Wilson, N.H.M. (Eds.), Proceedings of the Tenth International Symposium on Transportation and Traffic Theory. Elsevier, MIT, Cambridge, Massachusetts, pp. 61–68.

Smith, M.J., van Vuren, T., 1993. Traffic equilibrium with responsive traffic control. Transp. Sci. 27 (2), 118–132.

Smith, M.J., 2010. Intelligent network control: using an assignment - control model to design fixed time signal timings. In: Tampere, C., Viti, F., Immers, L. (Eds.), Transport Planning – Advances in Dynamic Traffic Assignment. Edward Elgar, pp. 57–72.

Smith, M.J., Huang, W., Viti, F., Tampere, C.T.M., Lo, H.K., 2019a. Quasi-dynamic traffic assignment with spatial queueing, control and blocking back. Transp. Res. Part B 122, 140–166.

Smith, M.J., Iryo, T., Mounce, R., Rinaldi, M., Viti, F., 2019b. Traffic control which maximises network throughput: some simple examples. Transp. Res. Part C 107, 211–228.

Smith, M.J., Iryo, T., Mounce, R., Satsukawa, K., Watling, D., 2022. Zero-queue traffic control, using green-times and prices together. Transp. Res. Part C 138, 103630.

Taale, H., van Zuylen, H.J., 2001. The combined traffic assignment and control problem – an overview of 25 years of research. In: *Proceedings of the 9th World Confernce on Transport Research* (WCTR), Seoul, Korea, pp. 22–27.

Tassiulas, L., Ephremides, A., 1992. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. IEEE Trans. Automat. Control 37 (12), 1936–1948.

Thompson, W.A., Payne, H.J., 1975. Traffic assignment on a transportation network with capacity constraints and queueing. In: Paper presented at the 47th National ORSA/TIMS North American Meeting, 1975.

Van Vuren, T., Van Vliet, D., 1992. Route Choice and Signal Control. Avebury.

Varaiya, P., 2013a. Max pressure control of a network of signalised intersections. Transp. Res. Part C 36, 177–195.

Varaiya, P., 2013b. The maxpressure controller for arbitrary networks of signalised intersections. In: Advances in Dynamic Network Modeling in Complex Transportation Systems, 57. Springer, pp. 27–66.
Viti and Van Zuylen, 2010. Probabilistic models for queues at fixed control signals. Transp. Res.Part B 44 (1), 172–184.
Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. In: Proceedings of the Institute of Civil Engineers Part II, 1, pp. 325–378.
Webster, F.V., 1958. Traffic Signal Settings. HMSO, London. Department of Transport Road Research Technical Paper 39.
Wongpiromsarn, T., Uthaicharoenpong, T., Wang, Y., Frazzoli, E., Wang, D., 2012. Distributed traffic signal control for maximum network throughput. In: Proceedings of the 15th International IEEE Conference on Intelligent Transport Systems.
Zhang, H.M., Yu (Marco), Nie, Zhen (Sean), Qian, 2013. Modelling network flow with and without link interactions: the cases of point queue, spatial queue and cell transmission model. Transp. B 1 (1), 33–51. https://doi.org/10.1080/21680566.2013.785921.