Available online at www.sciencedirect.com

SCIENCE DIRECT®

Transportation Research Part B 40 (2006) 731–744

TRANSPORTATION RESEARCH PART B

www.elsevier.com/locate/trb

Bus lanes with intermittent priority: Strategy formulae and an evaluation

Michael Eichler, Carlos F. Daganzo *

Department of Civil and Environmental Engineering, Institute of Transportation Studies, University of California, 109 McLaughlin Hall #1720, Berkeley, CA 94720, United States

Received 2 August 2005; received in revised form 30 September 2005; accepted 3 October 2005

Abstract

This paper evaluates strategies for operating buses on signal-controlled arterials using special lanes that are made intermittently available to general traffic. The advantage of special bus lanes, intermittent or dedicated, is that they free buses from traffic interference; the disadvantage is that they disrupt traffic.

We find that bus lanes with intermittent priority (BLIPs), unlike dedicated ones, do not significantly reduce street capacity. Intermittence, however, increases the average traffic density at which the demand is served, and as a result increases traffic delay. These delays are more than offset by the benefits to bus passengers as long as traffic demand does not exceed by much the maximum flow possible on the non-special lanes; the smaller the excess the better. BLIPs are not intended for roadways nearing or in excess of capacity.

The main factors determining whether an intermittent system saves time are: the traffic saturation level; the bus frequency; the improvement in bus travel time achieved by the special lane; and the ratio of bus and car occupant flows. In some scenarios where a dedicated bus lane could not be operated, a BLIP can save to bus and car occupants together as much as 20 persons-min of travel per bus-km. The required conditions for this to happen are quite particular. Typical savings are smaller. Formulae are given.

Keywords: Bus lanes; Intermittent bus priority; Transit signal priority; Intelligent transportation systems; Advanced public transit systems

1. Introduction

Urban traffic congestion severely impairs the effectiveness and attractiveness of bus systems. As a result, and despite their limited resources, transit agencies have to spend a considerable amount of time and effort implementing workarounds to the problem. Inexpensive solutions that do not involve new infrastructure are the most desirable.

* Corresponding author. Tel.: +1 510 642 3853; fax: +1 510 643 3955.
E-mail address: daganzo@ce.berkeley.edu (C.F. Daganzo).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2005.10.001

One low-cost option is transit signal priority (TSP). With TSP, buses can extend the green phase of traffic signals to claim the right-of-way and proceed unimpeded through an intersection. A handful of studies have documented the benefits of TSP implementations (Balke et al., 2000; Banerjee, 2001; Cima et al., 2000; Duerr, 2000; Furth and Muller, 2000; Garrow and Machemehl, 1998; Hunter-Zaworski et al., 1995; Janos and Furth, 2002; Kloos et al., 1995; Lin, 2002; Nash and Sylvia, 2001; Skabardonis, 2000). Unfortunately, TSP loses effectiveness with heavy traffic because the signals have to accommodate, not just the bus, but also the traffic in which it is embedded.

Dedicated bus lanes (DBLs) are another option. They may be combined with TSP to increase their impact. Unfortunately, DBLs remove one lane from general use and therefore reduce capacity. Obviously, DBLs are only appropriate for low traffic flows. This limitation can be partly overcome by opening the bus lane to general traffic intermittently when not in use by a bus.

Viegas and Lu (2001, 2004) seem to have been first in proposing and analyzing the concept of an intermittent bus lane (IBL). The system in these references restricts automobiles from changing into the bus lane ahead of the bus, but does not request those vehicles already there to leave the lane. As discussed in the references, signal adjustments would be used to flush the queues at traffic signals and clear the way for the bus. These signal adjustments may increase the amount of green time allocated to the arterial at times when the arterial demand is low, bus this could reduce the capacity of side streets and increase delay.

The IBL variant proposed here, termed "BLIP" (which is short for bus lanes with intermittent priority), forces traffic out of the lane reserved for the bus with variable message signs (VMS). BLIPs do not require changes to the signal settings. Therefore, they should be efficient and easy to evaluate. BLIPs can be combined with TSP, if desired.

This paper uses deterministic analysis techniques of traffic flow (kinematic wave) theory to study the feasibility, costs and benefits of BLIPs. We recognize that IBLs and BLIPs will not eliminate any problems currently experienced with DBLs, such as accommodating right turns and dealing with pedestrian interference. If these problems are pressing, infrastructure-intensive solutions such as bus-rapid-transit (BRT) may be required. Comparisons to BRT are not in the scope of our analysis. Therefore, BLIPs will only be compared to the DBL and "do-nothing" (mixed-traffic operation) alternatives. Section 2, evaluates the automobile carrying capacity of BLIP systems, and Section 3 shows how to estimate the travel time savings to both the automobile and bus occupants of an under-saturated BLIP system. Section 4 discusses the results and describes the proper domain of application for BLIPs.

2. Capacity analysis

A BLIP is essentially a set of rolling spatial cocoons (bus-lane sections) in which buses travel to the exclusion of other traffic. Each cocoon starts at the rear bumper of its bus and extends a fixed distance ahead. This zone is kept clear of non-bus traffic to ensure that the bus does not experience any delay. For practical reasons, the exclusion zone is assumed not to travel continuously along the roadway, but to advance discretely one block at a time. VMSs, possibly combined with in-pavement lights, would announce the changes. These changes would create temporary bottlenecks at the locations where lanes are dropped. These bottlenecks are the critical BLIP feature that an analysis should dissect.

Shaheen et al. (2005) presents a preliminary analysis of BLIPs, which uses a number of assumptions about cycle lengths, signal offsets and level of service constraints. They limit its generality. Our approach will yield rougher but simpler results and more general insights.

The proposed approach pertains to large systems, with so many blocks and bus stops that the street on which the bus moves can be treated as a homogeneous road without signals—i.e., where the disturbances of the traffic signals can be averaged in time and space. Buses are then modeled as slow vehicles that interfere with the flow of traffic, as "moving bottlenecks". The presence of buses reduces capacity and creates delay, but not as much as if a lane had been dedicated to the bus. This macroscopic idealization will reveal the main factors affecting performance, and simple formulae quantifying their effects. The next three subsections introduce supporting concepts and notation from kinematic wave theory (Section 2.1); describe the operation in more detail (Section 2.2); and estimate capacity (Section 2.3).

2.1. Kinematic wave theory

This analysis uses concepts of the kinematic wave (KW) theory proposed by Lighthill and Whitham (1955) and Richards (1956). This theory provides tested techniques for modeling traffic flow and queuing. It informs us on phenomena describing the dynamics of queue growth and discharge, the formation of stationary states in space–time, traffic response to signals and moving bottlenecks. The theory has limitations but can predict reasonably well average trip times over long distances, which is the metric of interest in our analysis.

One component of KW theory is a fundamental diagram (FD) that describes the relation between flow and density in the steady state. Our analysis assumes a triangular FD for all lanes combined, as displayed in Fig. 1. This is both simple and experimentally justified.

The flow at any given point on the diagram (a stationary "traffic state") will be expressed as a "q" with a subscript matching the label of the point on the diagram. For example, the flow at point E will be q_E. Fig. 1 depicts two curves. The outer, larger curve pertains to the full roadway, when all lanes are open to traffic. The inner, small curve describes the "reduced" roadway—when one of the lanes has been reserved for the bus and is therefore no longer available to private vehicles. Fig. 1 displays the following traffic states, which will turn out to be of interest throughout this paper:

O	empty roadway
A	generic uncongested
J	full roadway, jam density
C	full roadway, capacity
F	reduced roadway, jam density
E	reduced roadway, capacity
B	full roadway, congested conditions with same flow as state E
G	reduced roadway, congested conditions with same speed as B

Kinematic wave theory describes how a road in any initial condition (characterized by a distribution of states along its length), and with any feasible input flow, evolves over time. Of particular relevance for this paper are the moving bottleneck models in Gazis and Herman (1992) and Newell (1993, 1998).

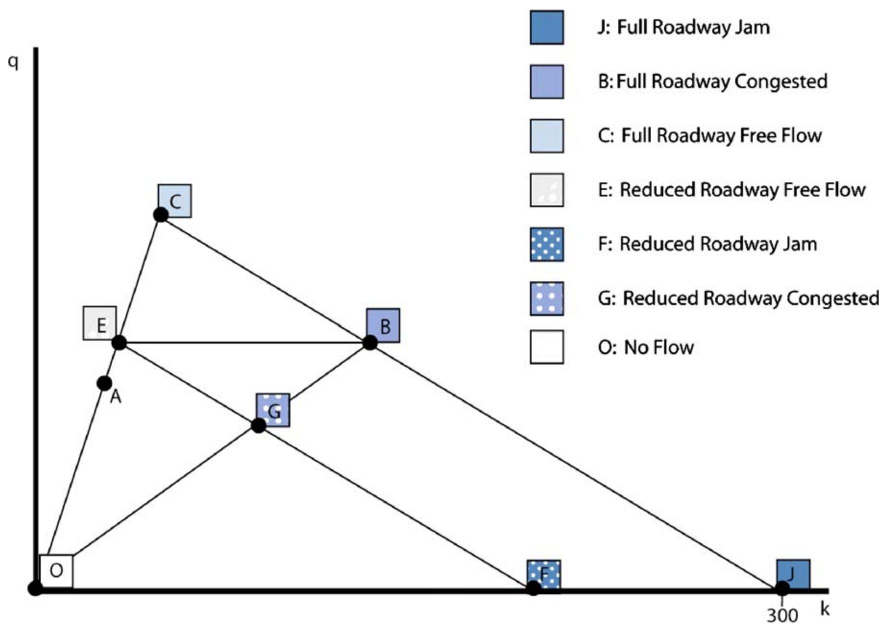

Fig. 1. Fundamental diagram illustrating full and reduced roadways.

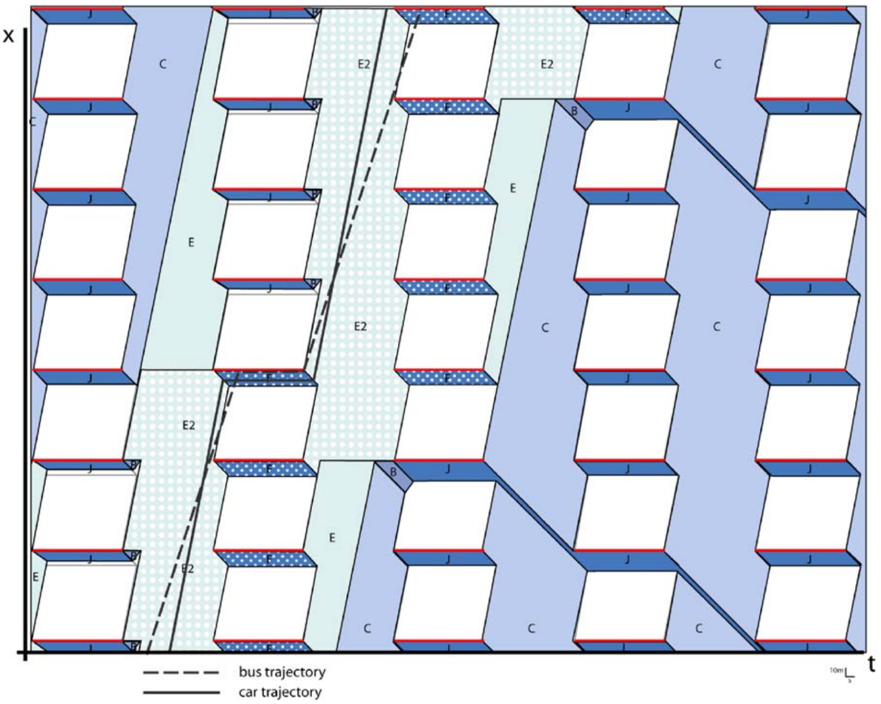

Fig. 2. Time–space diagram of a BLIP operating at capacity: traffic moves at maximum speed (faster than the bus) in all regions, except for F and J, where it is stopped, and B, where it moves slowly.

2.2. The BLIP operation and its effect on traffic

Fig. 2 is a time–space diagram constructed with KW theory of a BLIP system operating at capacity—with maximum entering flow from upstream. The diagram in Fig. 2 follows the convention for labeling traffic states described above. The states displayed in dotted shades of grey (E_2 and F) are restricted, i.e., correspond to the inner diagram with one less lane. The states displayed with solid shades are unrestricted. Traffic in restricted states cannot delay a bus. The diagrams assume that the restriction is announced by variable message sign (VMS) postings at every intersection. The VMS is activated at the far side of an intersection when it is determined that the vehicles discharging from that intersection would otherwise queue in front of the bus. These postings create a space–time region of (dotted) restricted states that allows the bus to travel unhindered by traffic—and vice versa. A dashed line depicts a hypothetical bus trajectory: the slope of this trajectory represents the average speed of the bus, factoring in delay from bus stops. The figure also depicts a car trajectory. Notice how the bus is allowed to pass the vehicles queued at the signal in state F, and how traffic is allowed to pass the bus, when in state E_2. Also notice that the bus passes through several intersections with green signals: the bus does not derive benefit from BLIP when passing green signals. The primary benefit to the bus is jumping traffic queues at intersections.

Our diagram assumes: (i) that postings last for a full cycle,[1] and (ii) that vehicles to have seen a posted restriction must obey it for the full length of the block until passing the VMS at the next intersection. Although the figure does not explicitly show the postings, the reader can verify that it is based on (i) from the result: note that the restriction must obviously be "on" at an intersection whenever its "world line" (the trajectory of the intersection) touches from below a restricted state E_2 during a green period; and "off" when the downstream state is unrestricted.

[1] Shaheen et al. (2005) also examine a less restrictive approach where postings can change in mid-phase, but find that the benefit of this generalization is small. Since the generalized approach is more complicated to implement, and could create confusion among drivers, it is not considered here.

As noted above, changeable message signs at the far side of every intersection communicate to drivers the BLIP status: lane restricted or lane available. This mechanism automatically flushes all vehicles out of the bus lane at every intersection immediately after its downstream restriction is activated—the vehicle queue on all lanes (state J) simply discharges into a reduced set of lanes (state E2) during the green period.[2] This is shown clearly by Fig. 2, which has been drawn to scale. In every such instance there is a small triangle of state B upstream of the intersection, which describes the congested conditions caused by the lane-drop bottleneck. These conditions have the effect of extending the time that it takes to discharge the queue, as the figure shows. Note that several VMSs may be activated simultaneously to clear enough space ahead of the bus.

Fig. 2 shows how disturbances grow and propagate when upstream traffic demand is large. The reduced queue discharge rates caused by VMS signs create queues of state J traffic at some signals that do not completely dissipate. Remnants propagate back to the upstream boundary, and manifest themselves as spillovers that block access.

2.3. Analysis of BLIP capacity

As the above analysis has indicated, a BLIP creates long-lasting queues that propagate upstream when traffic demand is at capacity—or close to capacity. Considering that subsequent buses can be delayed by these queues, further analysis is necessary to determine the collective effect of all the buses.

To this end, we consider a long bus route with many stops operated on a homogeneous road with many signals. The signals run with the same cycle, c, and green phases, g, but arbitrary offsets. (More generality would cloud the issues at hand; therefore, extensions will be discussed at the end of the paper.) We zoom out to a large scale of analysis, where the impacts caused by the signals can be averaged out and the street treated as if it was roughly homogeneous.

2.3.1. Macroscopic methodology

The fictitious road should have a set of stationary states that closely match the spatially and temporally averaged states of the real road. We assume that the intersections are sufficiently separated to guarantee that the capacity of the system is unaffected by the offsets. This is reasonable in any setting where a BLIP may be considered. The system capacity (maximum flow with signals) is then $q_{\mathrm{M}} = q_{\mathrm{C}}g/c$.

Our analysis assumes that the set of macroscopic states that can arise and be sustained on this street can be approximated by a trapezoid with corners at points O (the origin), M, N and the jammed state, J; see Fig. 4(a). The reason for this shape is that roadways with signalized intersections can support a maximum average flow (averaged over multiple cycles) at a range of average densities (averaged over multiple blocks). The lowest average density state that can support maximum flow (defined here as macroscopic state "M") occurs when an upstream queue discharges into an empty roadway. Using the traffic state convention described above, Fig. 3(a) illustrates the details of macroscopic state M for a pair of intersections with no offset. At the other end of the spectrum, the highest average density state that supports maximum flow (defined here as macroscopic state "N") should occur when a congested system dissipates from downstream: this is illustrated by Fig. 3(b). Note how the total outflow of this more congested system is the same as in Fig. 3(a). Because signalized roadways often exhibit a range (M to N) of traffic states with peak flow,[3] the FD used for a macroscopic analysis is trapezoidal.

This is reasonable. Patterns with average densities between k_{M} and k_{N} arise for example if parts of a street are in state M and other parts in state N; this dichotomous state of affairs is stable and sustainable. Hence, the

[2] To simplify Fig. 2, we have assumed that the VMS restriction is placed at the intersection threshold and that this placement would not change the queue discharge rate per (available) lane. This is slightly unrealistic. Maximum discharge rates, however, can be strictly preserved by moving the sign farther downstream; e.g., to the middle of the block. This is also a better solution from a human factors perspective. We did not present it because it yields more complicated diagrams. Fortunately, as we shall soon see, the macroscopic analysis of this paper does not depend on the specific location of the sign; only on the discharge rates being fixed. Therefore, our results apply to the favored placement even more accurately than to the threshold placement.

[3] The relative position of points M and N depends in a complicated way on the timing plan. For some plans, M = N; for others they are far apart. See Newell (1981) for more discussion.

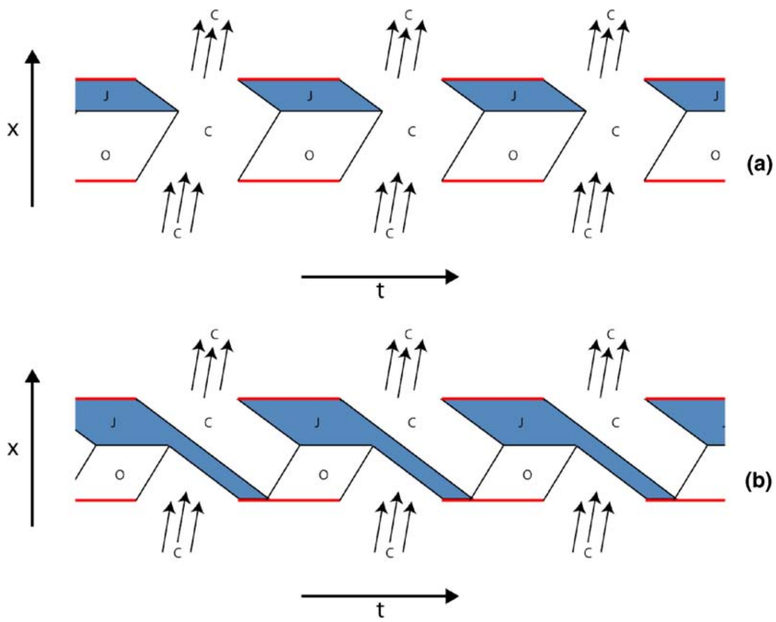

Fig. 3. Time–space diagram of an arterial operating at capacity: (a) light average density; (b) heavy average density.

horizontal line containing points M and N will be called the capacity line. The slope of the rising (left) branch of the trapezoid, denoted u, is the average speed of traffic in state M. Linearity implies that this speed is assumed to be the same for all uncongested states, although in reality the average traffic speed declines with flow. To capture this effect we introduce a separate parameter for the speed when the road is empty, v_0, which should be greater than u. (A better approximation would recognize that the rising branch of the trapezoid is concave, since speed declines with flow, but the shape of this branch depends on details of the signal settings that are cumbersome to specify. The effort would be of little value because the uncongested states that arise in our analysis are usually close to M.) We also use a linear dropping branch, since its shape is even less important. Note that the slope of this branch must be shallower than the slope of the FD for the road without signals. There is a reason for this: since the coordinates of point N are generalized averages in the sense of Edie (1963), point N must be at the center of gravity of points O, C and J, when they are weighted by their areas on the time–space plane; obviously, point N must be interior to triangle OCJ.

By definition, the new FD matches the stationary states observed on the real road. We propose that this FD can also be used with KW theory to describe the macroscopic dynamics of the road with and without BLIPs. To see that this is reasonable let us examine the backward wave speed of the modified road. The backward wave speed is an important determinant of dynamic behavior, since it is the speed at which disturbances propagate inside queues, which helps determine their length. Note from Fig. 4(a) that the wave speed of the new FD (the slope of the dropping branch) is significantly less than the original, as we have already discussed. Reassuringly, we see that the disturbances "B" of Fig. 2 are delayed at the traffic signals; they travel with the original wave speed between signals, but experience delays and indeed travel with a lower average speed. The reader can verify that their average speed is indeed the wave speed of the modified FD. Thus, we can be confident that treating the road as we are suggesting is reasonable, even in the dynamic case. This is convenient because buses and their cocoons can then be modeled as KW moving bottlenecks.

2.3.2. Effect of a single bus on a long street

For ease of explanation we initially treat these bottlenecks as points (neglecting the spatial extent of the cocoons) and generalize the results later. Moving bottlenecks can create different traffic conditions upstream and downstream of their locations—when they hold back a queue. When this happens, upstream traffic is in a congested state (U) and downstream traffic in a freely flowing state (D). The flow of state D (the bottleneck capacity) is assumed to be the capacity of the reduced system (minus one lane) including the effect of signals,

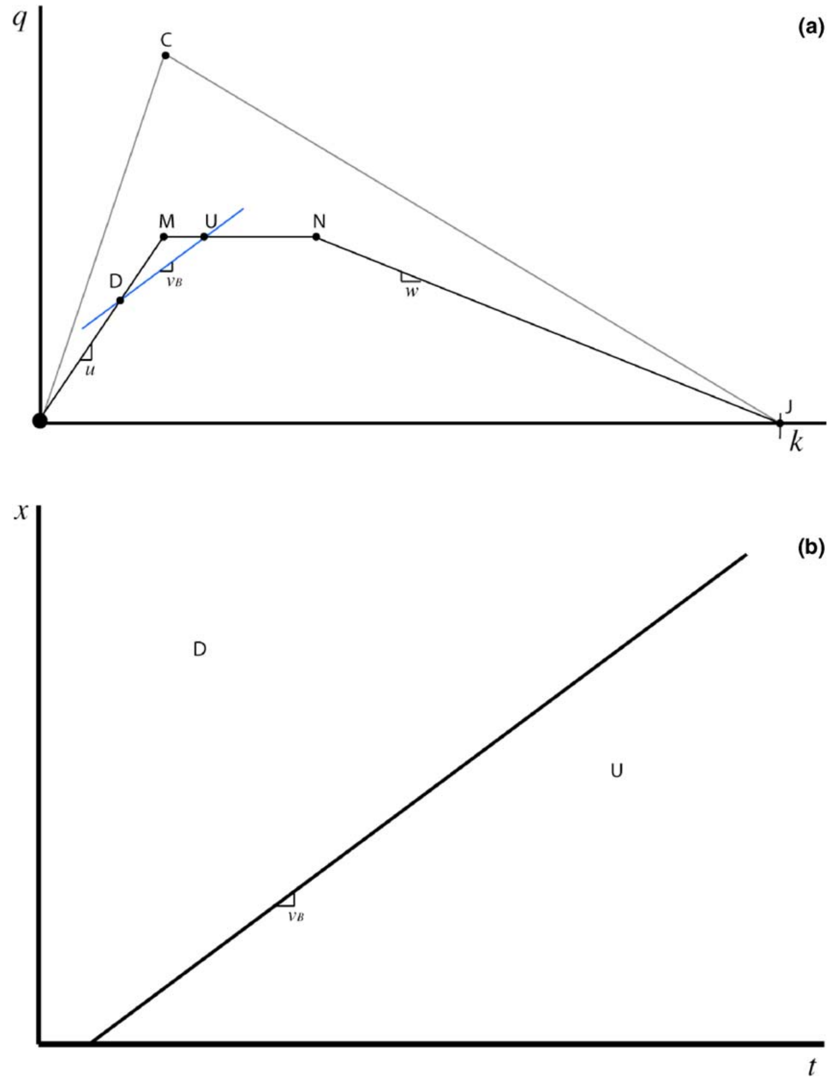

Fig. 4. Macroscopic diagrams of a BLIP system: (a) fundamental diagram showing moving bottleneck; (b) time–space diagram showing effect of a single bus.

$q_D = q_M(n-1)/n$, in agreement with Gazis and Herman (1992) and Newell (1993, 1998). Conservation of flow past the bottleneck implies that the interface between states D and U (the bus), which travels at an average speed of v_B, is expressed in the FD by a line with slope v_B from state D to state U. Thus, the position of state U on the FD is completely determined; see Fig. 4(a). Note that state U has higher flow than D. If the speed of the bus is sufficiently high, as occurs in the figure, $q_U = q_M$. For typical values of the parameters, even if U is on the dropping branch, $q_U \approx q_M$.

According to KW theory, if the road was infinitely long and there was an infinite demand waiting to enter, the introduction of a single BLIP bus would result in the pattern of Fig. 4(b) in the neighborhood of the bus. If the dimensions of the bus cocoon are non-zero but too small to be discernible in the picture, it is obvious that the states illustrated would be the same: the size of the cocoon is trivial compared to the sizes of the neighboring traffic states. (Additionally, traffic state D is, in effect, the cocoon.) Thus, the figure applies to cocoons of any size when used to examine the evolution of the system for an indefinitely long time. We see that the beginning of the roadway would switch from state D to state U after the bus passage. Hence, q_U is the maximum flow that could be sustained; i.e., the capacity (q_{max}) of the single-bus BLIP system on an infinitely long street. Thus, $q_{\max} = q_U \approx q_M$. This shows that the introduction of a single BLIP bus does not significantly reduce the street capacity, which was q_M to begin with. By comparison, a dedicated bus lane reduces street capacity by $100/n\%$ from q_M to q_D. This suggests that BLIPs should have the most to offer when traffic demand exceeds q_D because then a dedicated lane is infeasible.

2.3.3. Multiple buses

Extending the analysis to more than one bus is easy. Fig. 5 illustrates the situation where the BLIP lane makes up a portion of length L of the roadway in question and buses follow each other with headway H. We assume that buses are not coordinated with the signals and (momentarily) that the time-dimension of the cocoon (which is comparable with the signal cycle c) is small compared with the headway (i.e., $c \ll H$.) The fundamental diagram for this situation, Fig. 5(a), indicates that the traffic demand is in a state A with flow $q_A \in [q_D, q_U]$. The time–space solution in Fig. 5(b) shows that the state introduced by each bus downstream of itself, D, meets with the congested upstream traffic state U from the previous bus, canceling out a wedge of state A. The wedge may be truncated if the road is very short. In either case, the average flow across any point on the road is q_A, independently of L and H, and this flow can be sustained for any number of headways. Thus, any flow $q_A \leqslant q_U$ can be sustained. We also see from the analysis that demands q_A greater than q_U cannot be accommodated. Thus, q_U ($\approx q_M$) continues to be the car-carrying capacity of the system, provided $H \gg c$.

For smaller H, the scale of Fig. 5(b) would be such that the size of the cocoon would become nontrivial and should be included in the analysis. It can be seen in Fig. 2 that the actual time-width of the cocoon is about two cycle lengths. If we define $\hat{c}$ as the cocoon width ($\hat{c} = 2c$), the bus trajectories of Fig. 5(b) would become

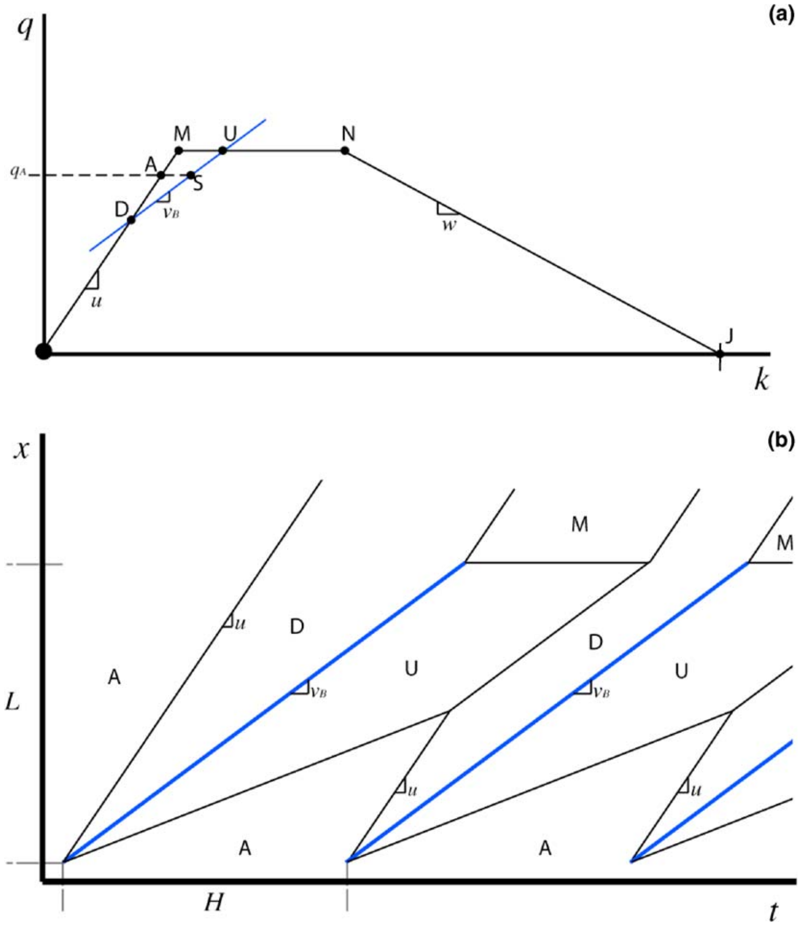

Fig. 5. Macroscopic diagrams for a multi-bus BLIP system: (a) fundamental diagram; (b) time–space diagram with multiple buses.

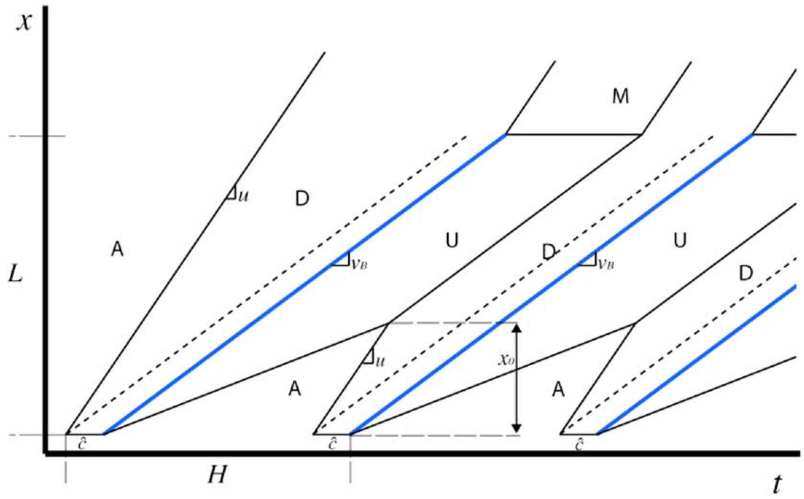

Fig. 6. Time–space diagram illustrating situation when the cocoon time-width is accommodated. The dashed line represents the extent of the bus cocoon.

bands of time-width $\hat{c}$ in which state D would prevail. As shown in Fig. 6, if H $\geqslant \hat{c}$ the geometric pattern between the bands is similar to the original, of time duration $H - \hat{c}$ instead of H, and the maximum possible demand in the inter-band portions would continue to be q_U. By prorating the maximum flows in the band and inter-band regions we obtain the following formula for the automobile-carrying capacity of the system:

$$q_{\max} = q_D(\hat{c}/H) + q_U(1 - \hat{c}/H) \quad \text{for } H \geqslant \hat{c}. \tag{1}$$

We assumed in the construction of Fig. 5 that state U is on the capacity line. If the bus speed is so low that U is on the declining branch of the FD, the diagram of Fig. 5(b) would look only slightly different. A picture would reveal that in this case too, a flow equal to q_U can always be accommodated, and that $q_U \approx q_M$, as we have already mentioned. If we make this substitution in (1), and also use the relation $q_D = q_M(n-1)/n$, we find

$$q_{\max} \approx [1 - \hat{c}/(nH)]q_M \quad \text{for } H \geqslant \hat{c}. \tag{2}$$

The ratio $\hat{c}/(nH)$ appearing in this formula is the fractional reduction in a street's car-carrying capacity caused by the BLIP. Values on the order of just a few percent should be typical—the reduction is 5% if $n = 4$, $c = 1$ min, and $H = 10$ min.

To summarize, we have shown that a BLIP can accommodate a car flow up to a level close to q_M independently of the BLIPs length if the street has three lanes or more and the BLIP headways are considerably longer than the signal cycle. Under these conditions, BLIPs do not significantly reduce road capacity. However, as should be clear from Fig. 4, BLIPs do change the *character* of capacity traffic by increasing the average density at which capacity is achieved. This change in character affects level of service, and this is examined next.

3. Level of service analysis

We consider in this section the delay imposed by a BLIP on a traffic stream, assuming that the demand does not exceed capacity. We will express the results in terms of "average pace"; that is the average number of minutes required to travel a mile. We will do this, first for the traffic stream and then for the bus. We will compare the change in pace for both modes before and after the BLIP.

3.1. The effect of BLIP on automobile delay

We first consider the asymptotic case of a very long bus line, with $L \to \infty$, and then examine the corrections due to end effects. We assume that the bus average speed is such that $q_U \approx q_M$ and that the demand is in the range $q_A \in [q_D, q_U]$.

3.1.1. Long roads

Assume to begin with that $H \gg c$ and neglect the width of the cocoon. The diagram of Fig. 5(b) applies. If for a given FD, L is allowed to grow while H is held constant, the wedge of state A stays pinned to the upstream end of the diagram without any change, but the interface between states U and D will grow. For very large L, the diagram consists of alternating parallel bands of states U and D, with a negligible wedge of state A at the bottom. We assume here that the bands extend from the beginning to the end of the road. The correction due to the wedges will be presented later; it is negligible for large L.

Let H_{U} and H_{D} denote the time span of each band, such that

$$H_{\mathrm{U}} + H_{\mathrm{D}} = H. \tag{3a}$$

Since the average flow in one headway must be q_{A} at all locations, we have

$$q_{\mathrm{U}}H_{\mathrm{U}} + q_{\mathrm{D}}H_{\mathrm{D}} = q_{\mathrm{A}}H, \quad \text{where } 0 \leqslant q_{\mathrm{D}} \leqslant q_{\mathrm{A}} \leqslant q_{\mathrm{U}} = q_{\mathrm{M}}. \tag{3b}$$

This determines the relative width of each band.

The total number of vehicle-minutes spent by drivers between two consecutive buses can be written using Edie's generalized (average) density for the time–space region between buses, which is

$$k_{\mathrm{S}} = (k_{\mathrm{U}}H_{\mathrm{U}} + k_{\mathrm{D}}H_{\mathrm{D}})/H. \tag{4a}$$

The expression for vehicle-minutes of travel per bus is (see e.g., Daganzo, 1997),

$$\text{Total time per bus} = LHk_{\mathrm{S}}. \tag{4b}$$

Recall now that Edie's generalized flow for the time–space region between buses is $q_{\mathrm{S}} = (q_{\mathrm{U}}H_{\mathrm{U}} + q_{\mathrm{D}}H_{\mathrm{D}})/H \equiv q_{\mathrm{A}}$. Thus, the generalized average state between buses is point $(k_{\mathrm{S}}, q_{\mathrm{A}})$ of the density-flow plane. As shown by Fig. 5(a), this is the point at which the horizontal line for flow q_{A} intersects segment DU. The speed associated with point S is the average speed of traffic. The length of segment AS, denoted |AS|, is proportional to the increase in travel time caused by the BLIP. Fig. 5(a) shows at a glance that this penalty increases linearly with both, the demand, q_{A}, and the average bus pace, $1/v_{\mathrm{B}}$. In fact, letting Δ denote the increase in the vehicle-minutes of automobile travel induced by one bus-kilometer we have

$$\Delta = |AS|H = (k_{\mathrm{S}} - k_{\mathrm{A}})H = (q_{\mathrm{A}} - q_{\mathrm{D}})(1/v_{\mathrm{B}} - 1/u)H \quad \text{for } H \gg c. \tag{5a}$$

If the physical dimension of the cocoon ($\hat{c} = 2c$) is incorporated into the analysis, bands of state D replace the bus trajectories of Fig. 5(b) resulting in a picture such as Fig. 6, and the overall portion of the diagram covered by state D increases. Because traffic in state D travels at free-flow speed, the overall average traffic speed in the time–space region between bus headways *increases*, and (5a) is an upper bound to the penalty imposed by a BLIP. The actual penalty can be easily derived by repeating the analysis; it turns out to be $(1 - \hat{c}/H)$ times smaller; i.e.,

$$\Delta = (1 - \hat{c}/H)(q_{\mathrm{A}} - q_{\mathrm{D}})(1/v_{\mathrm{B}} - 1/u)H \quad \text{for } H \geqslant \hat{c}. \tag{5b}$$

The correction makes sense: if $\hat{c} = H$ the BLIP would behave like a dedicated lane, which does not delay traffic as long as demand stays below capacity (2)—just as predicted by (5b).

3.1.2. Short roads

If the road is short Eq. (5) overestimates the BLIP penalty because it assumes that the initial triangular wedge is in state S when it is actually in state A (with no delay). The spatial extent of the wedge, x_0 (see Fig. 6), is easily obtained from the slopes of its sides and the dimension of its base $(H - \hat{c})$. We find

$$x_0 = (1 - q_{\mathrm{A}}/q_{\mathrm{M}})n(H - \hat{c})/(1/v_{\mathrm{B}} - 1/u). \tag{6}$$

The vehicle-hour penalty in the first x_0 miles of road only applies to 1/2 of the band between buses; i.e., the part not covered by the wedge. Thus, to be precise, penalty (5) should be applied to a road that is shortened by $x_0/2$ distance units; i.e., the exact formula is

$$\text{Total VHT penalty per bus} = \Delta\left(L - \frac{1}{2}x_0\right), \quad \text{if } L \geqslant x_0. \tag{7a}$$

If $L \leqslant x_0$ the wedge reaches the downstream end of the road, and the penalty per bus is fixed, independent of the headway. This penalty should increase with the square of L/x_0 and equal the value of (5a) for $L = x_0$. Thus, we have

$$\text{Total VHT penalty per bus} = \frac{1}{2}\Delta L^2 x_0, \quad \text{if } L \leqslant x_0. \tag{7b}$$

3.2. The effect of BLIP on bus pace: cost–benefit comparisons

The bus performance results are clearer if expressed in terms of pace. Thus, we introduce $p = 1/u$, $p_0 = 1/v_0$ and p_f as the prevailing, traffic-free and signal-free automobile paces, respectively. The last parameter corresponds to the speed limit. Typical values for an arterial street are: $p = 1.9$, $p_0 = 1.7$ and $p_f = 1.3$ min/km. The bus paces should be roughly equal to the auto paces plus the bus stop-time per kilometer, τ. If we use the letter b with the same set of subscripts to denote bus paces (in mixed traffic, with a BLIP and with a BLIP/TSP), we have: $b \approx p + \tau$, $b_0 \approx p_0 + \tau$ and $b_f = p_f + \tau$. We actually expect b to be slightly greater than $(p + \tau)$ if the signal system is designed to accommodate the prevailing traffic without regard for the bus stops; perhaps on the order of 0.2 min/km.[4] If $\tau \approx 1$ min/km (a reasonable value for a line with infrequent stops) then the bus paces on our hypothetical arterial could well be: $b = 3.1$, $b_0 = 2.7$ and $b_f = 2.3$ min/km.

A BLIP implementation should reduce the bus pace from b to b_0; i.e., by about 0.4 min/km. If the number of passengers in the average bus is O (pax) then the passenger-minutes saved per bus-km traveled, δ, is

$$\delta = O(b - b_0) \quad (\text{pax-min/bus-km}). \tag{8}$$

If $O \approx 30$ pax then $\delta \approx 12$ pax-min/bus-km in our hypothetical scenario.

Society should decide the criteria for implementation of a BLIP, but a critical factor in this decision should be the relative magnitude of δ (capturing the benefit to transit patrons) and Δ (capturing the disbenefit to drivers). Assuming that automobile occupancies are close to 1, society should probably not consider BLIP if δ is much smaller than Δ. In terms of pace (5) indicates that $\Delta \leqslant (q_A - q_D)(b - p)H$. In our typical example, $(b - p) = 1.2$ min; then, if we take $H \approx 5$ min (at the lower end of the values reasonable for a BLIP) we find $\Delta \leqslant 6(q_A - q_D)$, where flows are expressed in veh/min. This value should be comparable or smaller than $\delta = 12$ (min) for a BLIP to be appealing; i.e., the traffic demand should satisfy $(q_A - q_D) < 12/6 = 2$ veh/min, or 120 veh/h. This is the value by which traffic demand for a BLIP can exceed the capacity of a system with a dedicated lane and still be of some benefit.[5]

Of course, this conclusion depends on the values of the parameters we have chosen for the comparison. A quick test for long roads can be based on the inequality $\delta/\Delta > 1$, which, after grouping terms of (5b) and (8), reduces to

$$[O/Hq_A][(b - b_0)/\tau] > [1 - q_D/q_A][1 - \hat{c}/H] \quad (\text{for long roads}). \tag{9}$$

The first term of (9) is the ratio of bus-passenger to car-flow (modal split); the second the ratio of bus travel time reduction to bus-stop delay (improvement in bus service); the third (on the right side) the fractional amount by which traffic flow exceeds the capacity of the arterial with a dedicated bus lane (traffic saturation level); and the fourth the fraction of traffic cycles unaffected by the BLIP (a dimensionless measure of bus frequency). If the inequality is not satisfied, it is most efficient in terms of people's cumulative time savings to operate the bus in mixed traffic. Note that the inequality is most likely to be satisfied when τ is small, suggesting that BLIPs should be most successful when used for express bus service.

The comparison we have made assumes that $q_A > q_D$ but BLIPS can also be considered if the demand is lower. In this case, the analysis methodology of this paper would show that introduction of a BLIP does not disrupt traffic significantly. But if demand is significantly lower than q_D (say 400 veh/h or less for $n = 3$ or 4), one should also be able to introduce a dedicated lane without much disruption. Therefore, a BLIP has a distinct advantage over other alternatives only if the demand is close to q_D; e.g., $q_A \in (q_D - 400,$

[4] If turning traffic interferes with bus performance significantly, b_0 should also be increased.

[5] In actuality, the limit should be slightly larger because we did not account for the benefit of added reliability.

$q_D + 120$) veh/h in our hypothetical scenario. This range of flows can be expanded if the BLIP is combined with TSP. Note that a BLIP operation can be turned off or morphed into a dedicated lane at will. Therefore, it may be useful where traffic conditions or bus frequencies change with time.

3.3. BLIP/TSP systems

The advantage of a BLIP/TSP operation is that it further reduces the bus pace to b_f. The benefit to bus passengers can be quantified with (8), which continues to apply with b_f substituted for b_0. The disadvantage of a BLIP/TSP operation is that it is more complex and potentially disruptive of automobile traffic. Fortunately, as we shall soon see, the increased disruption is usually small.

We assume that buses preempt signals by shortening the red phase in which they would otherwise arrive, and that they do so by the least amount necessary to receive a green phase. We also assume that buses arrive at a signal immediately downstream of a stop independently of its cycle—due to the random necessity for stopping. Under these assumptions, the probability of a bus arriving during the red phase (needing preemption) is r/c. To accommodate such arrivals, the red phase will on average have to be reduced by 1/4 of its length[6]—by terminating it earlier or starting it later; i.e., by $1/4r$. Thus, the unconditional expected reduction in red time per bus arrival should be: $(r/c)(1/4r) = 1/4r^2/c$.

To leave side streets as unaffected as possible, we further assume that the reduction in their green time due to the passage of a bus is canceled in ensuing cycles with an offsetting increase of the same magnitude. Thus, the arterial red time will increase in the headway following the passage of a bus by an amount averaging $1/4r^2/c$. The decrease in red time (increase in green time) occurs when the discharge rate is reduced by one lane, but the decreased green time occurs when it is not. Thus, there is a net loss of arterial capacity at the intersection. The loss equals:

$$\text{Capacity loss due to TSP} = [1/4r^2/cH][q_M/n]. \tag{10}$$

Fortunately, this is usually a small number. If we take (conservatively) $r/c = 1/2$ and $r/H = 0.1$ the first factor is 0.0125. But it should be smaller in most cases. Thus, we see that TSP, if implemented properly, imposes a capacity penalty roughly equivalent to (at most) 1% of the capacity of a single lane. This is insignificant. Furthermore, a (small) capacity reduction has no discernible effect on the approximate traffic analysis of Section 3.2. Thus, (5) continues to apply.

It follows that (9) can also be used to assess the suitability of a BLIP/TSP if b_f substituted for b_0. We find for the same data of Section 3.2 that $\delta = 24$ instead of 12 (pax-min/bus-km). Thus, the range of applicability is expanded to $(q_A - q_D) < 24/6 = 4$ veh/min, or 240 veh/h above the system capacity with a dedicated lane. We expect the competitive advantage of a BLIP/TSP over a dedicated lane with preemptive bus priority to decline quickly as in the case of a pure BLIP. Roughly speaking, a BLIP/TSP doubles the maximum excess demand where intermittence reduces time in the system, and also doubles the maximum possible reduction—from about 10 to about 20 pax-min/bus-km. Reductions of this magnitude, however, can only be expected when traffic demand is very close to the system capacity with a dedicated lane.

4. Discussion

We have examined in this paper the effects of BLIPs, with and without TSP, but have not commented on the benefits of signal priority without a dedicated lane. Pure TSP strategies are not real competitors with BLIP or BLIP/TSP when the latter can be used. If we were running a signal preemption system without reserving a lane, but the lane could be reserved without significant disruption to traffic, then converting the preemption system to a BLIP/TSP would yield some improvement since the bus would avoid all queues and the preemption times could be shortened. Therefore, we conclude that pure preemption should only be used when the

[6] If the signal priority scheme was either only red truncation or green extension, the expected reduction in red signal time would be $1/2r$: the expected value of a uniform random variable that ranges from 0 to r. However, the signal can be modified in two directions: either the green can be extended (for buses arriving in the first half of the red signal) or the red can be truncated (otherwise). Therefore, the average red phase reduction is the expected value of a uniform random variable ranging from 0 to$1/2r$:$1/4r$.

demand exceeds the upper limit of the BLIP/TSP range of applicability. Thus, we suggest the following (rough) domains of application for the transit management strategies in the scope of this paper:

(1) DBLs and DBL/TSP: demand less than 80% or 90% of q_D;
(2) BLIP or BLIP/TSP: demand close to q_D; and
(3) Pure TSP, with queue jump lanes if possible[7]: demand larger than 120% of q_D.

This ranking is mostly qualitative, but it shows that BLIPs have a definite niche in the ecosystem of bus-friendly transportation management strategies. Sharper boundaries can be defined with the formulae of this paper for specific system configurations. Eq. (9) for example shows that the benefits are likely to be most pronounced for express bus service (when τ is small). The formulas, however, are only approximations; they should be complemented with more detailed study if an implementation is being considered.

The formulas in this paper assume that the lane-changes created by the VMS signs do not reduce the saturation flow per lane at the signals. But this is optimistic if significant lane changes are allowed to occur near the signals. We have already mentioned that to avoid this problem, VMS restrictions should be put in place, not at the intersection threshold, but toward the middle of the block; this would in addition better accommodate right turn movements. Eichler (2005) discusses design issues in more detail.

This analysis treats many aspects of BLIPs as black boxes: the operational details of many of the technical components are not important to the potential costs and benefits of a BLIP. For example, it matters little whether the application logic is centralized or not. It is assumed, however, that the vehicle location system and application logic work together to maintain an accurate trajectory projection for the bus, and that this projected trajectory is the basis for control of the VMS. Eichler (2005) discusses the technology components that can be used to implement BLIPs.

The formulas of the paper also assume that the system is homogeneous, has little turning traffic and is time-independent. This is sufficient to derive some insights, but the formulae should be modified if these effects are important. We propose that the traffic dynamics of a BLIP, including all these complications, can be roughly described by the KW model with a truncated FD typified by the truncated trapezoid ODUNJ of Fig. 5(a). This is reasonable because the macroscopic steady states of the system (on a scale of observation large compared with the bus headway and the bus spacing) fall on the truncated trapezoid. Since the traffic stream is modeled as a homogeneous stream with no-passing, kinematic waves make sense. The BLIP FD has four wave speeds, and the reader can verify (with some effort) that indeed the transitions between regimes propagate with the required wave speeds. For example, transitions between states on branch DU can be shown always to be contained between two consecutive buses, and therefore to propagate with speed v_B, as required. The kinematic wave model allows one to examine a system with entering and exiting traffic very quickly and to analyze it numerically with very little effort. The results are, of course, only approximate, but can be obtained without the tedium and potential for large execution errors arising when setting up a micro-simulation.

We have not attempted in the above to quantify all the benefits of BLIP/TSP service. In addition to the estimated time reduction, BLIP/TSP can also reduce random fluctuations in travel and arrival times, which should further enhance the appeal of the service. These added benefits could perhaps be a deciding factor in cases where total user time in the system is not significantly changed.

Acknowledgement

Research supported by the U.C. Berkeley Center for Future Urban Transport (a Volvo Center of Excellence).

[7] A queue jump lane is a shoulder-side flare on the upstream side of an intersection that is reserved for buses and right-turning vehicles (Rosinbum et al., 1991; TRB, 2000; Mirabdal and Thesen, 2002). The extra lane allows buses to "jump" the traffic queues at the signal. These lanes often have special signalization that allows the bus to pull into the intersection before the vehicles in the other lanes, giving the bus priority as it returns to the through-traffic lane. This is attractive but BLIPs do the same without the expense of additional right-of-way.

References

Balke, K., Urbanik, D., Conrad, L., 2000. Development and evaluation of intelligent bus priority concept. Transportation Research Record 1727, 12–19.

Banerjee, F., 2001. Transit priority system evaluation report, City of Los Angeles Department of Transportation, April 2001.

Cima et al., 2000. Transit signal priority: a comparison of recent and future implementations, paper presented at ITE 2000 Annual Meeting, Nashville, TN, 6–10 August 2000.

Daganzo, C.F., 1997. Fundamentals of Transportation and Traffic Operations. Pergamon-Elsevier, New York, NY.

Duerr, P., 2000. Dynamic right-of-way for transit vehicles: integrated modeling approach for optimizing signal control on mixed traffic arterials. Transportation Research Record 1731, 31–39.

Edie, L.C., 1963. Discussion of traffic stream measurements and definitions. In: Almond, J. (Ed.), Proceedings of the 2nd International Symposium On the Theory of Traffic Flow. OECD, Paris, France, pp. 139–154.

Eichler, M.E., 2005. Bus lanes with intermittent priority: assessment and design masters of city planning thesis, Department of City and Regional Planning, University of California, Berkeley.

Furth, P., Muller, T., 2000. Conditional bus priority at signalized intersections: better service with less traffic disruption. Transportation Research Record 1731, 23–30.

Garrow, M., Machemehl, R., 1998. Development and evaluation of transit signal priority strategies. Center for Transportation Research, University of Texas, Austin, Washington, DC, Transportation Research Board.

Gazis, D.C., Herman, R., 1992. The moving and 'phantom' bottlenecks. Transportation Science 26, 223–229.

Hunter-Zaworski, K., Kloos, W., Danaher, A., 1995. Bus priority at traffic signals in Portland: the Powell Boulevard pilot project. Transportation Research Record 1503, 29–33.

Janos, M., Furth, P., 2002. Bus priority with highly interruptible traffic signal control: simulation of San Juan's Avenida Ponce de Leon. Transportation Research Record 1811, 157–165.

Kloos, W., Danaher, A., Hunter-Zaworski, K., 1995. Bus priority at traffic signals in Portland: the Powell Boulevard pilot project. In: Proceedings of the annual meeting of ITS America, vol. 1, pp. 499–503.

Lighthill, M., Whitham, J., 1955. On kinematic waves. I. Flow movement in long rivers. II. A theory of traffic flow on long crowded road. Proceedings of Royal Society A229, 281–345.

Lin, W., 2002. Quantifying delay reduction to buses with signal priority treatment in mixed-mode operations. Transportation Research Record 1811, 100–106.

Mirabdal, J., Thesen, B., 2002. Using video detection for transit priority. ITE Conference.

Nash, A., Sylvia, R., 2001. Implementation of Zürhch's Transit Priority Program. Mineta Transportation Institute, College of Business, San José State University, San Jose, CA, October.

Newell, G.F., 1981. Blocking effects for synchronized signals. In: Proceedings, 8th International Symposium on Transportation and Traffic Theory. University of Toronto Press, Toronto, Canada.

Newell, G.F., 1993. A moving bottleneck. Institute of Transportation Studies Research Report UCB-ITS-RR-93-3, University of California, Berkeley, California. (In abridged form, Trans. Res. 32B, 531–537, 1998).

Newell, G.F., 1998. A moving bottleneck. Transportation Research, Part B 32 (8), 531–537.

Richards, P.I., 1956. Shockwaves on the highway. Operations Research 4, 42–51.

Rosinbum, T., Grote, W., Jickling, D., 1991. Bus bay street-related improvements in Phoenix and Tucson. ITE Journal 61 (8), 19–22.

Shaheen, S., Eichler, M., Daganzo, C.F., Barth, M., Todd, M., 2005. Enhanced Transit Strategies: Bus lanes with Intermittent Priority and ITS Technology Architectures for TOD Enhancement PATH MOU 5103 Final Report, Institute of Transportation Studies, University of California, Berkeley, CA.

Skabardonis, A., 2000. Control strategies for transit priority. Transportation Research Record 1727, 20–26.

Transportation Research Board (TRB), 2000. Highway Capacity Manual (NRC).

Viegas, J., Lu, B., 2001. Widening the scope for bus priority with intermittent bus lanes. Transportation Planning and Technology 24 (2), 87–110.

Viegas, J., Lu, B., 2004. The intermittent bus lane signals setting within an area. Transportation Research, Part C 12, 453–469.