



# A systematic analysis of multimodal transport systems with road space distribution and responsive bus service



Fangni Zhang<sup>a,b,c</sup>, Nan Zheng<sup>d,c,\*</sup>, Hai Yang<sup>b</sup>, Nikolas Geroliminis<sup>c</sup>

<sup>a</sup> School of Aviation, Faculty of Science, University of New South Wales, Sydney, Australia

<sup>b</sup> Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China

<sup>c</sup> Urban Transport Systems Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

<sup>d</sup> Institute of Transport Studies, Department of Civil Engineering, Monash University, Australia

## ARTICLE INFO

### Keywords:

Multimodal system  
Dedicated bus lane  
Responsive transit service  
Macroscopic fundamental diagram  
Dynamic user equilibrium

## ABSTRACT

A smart design of transport systems involves efficient use and allocation of the limited urban road capacity in the multimodal environment. This paper intends to understand the system-wide effect of dividing the road space to the private and public transport modes and how the public transport service provider responds to the space changes. To this end, the bimodal dynamic user equilibrium is formulated for separated road space. The Macroscopic Fundamental Diagram (MFD) model is employed to depict the dynamics of the automobile traffic for its state-dependent feature, its inclusion of hypercongestion, and its advantage of capturing network topology. The delay of a bus trip depends on the running speed which is in turn affected by bus lane capacity and ridership. Within the proposed bimodal framework, the steady-state equilibrium traffic characteristics and the optimal bus fare and service frequency are analytically derived. The counter-intuitive properties of traffic condition, modal split, and behavior of bus operator in the hypercongestion are identified. To understand the interaction between the transport authority (for system benefit maximization) and the bus operator (for its own benefit maximization), we examine how the bus operator responds to space changes and how the system benefit is influenced with the road space allocation. With responsive bus service, the condition, under which expanding bus lane capacity is beneficial to the system as a whole, has been analytically established. Then the model is applied to the dynamic framework where the space allocation changes with varying demand and demand-responsive bus service. We compare the optimal bus services under different economic objectives, evaluate the system performance of the bimodal network, and explore the dynamic space allocation strategy for the sake of social welfare maximization.

## 1. Introduction

City centers nowadays are experiencing significant amount of daily commuting, causing traffic congestion to notoriously spread in urban road networks. Constructing new infrastructure is expensive, while implementing road pricing is practically difficult due to user acceptability. Considerable efforts have been devoted to smart planning and allocation of the existing road space among different modes, e.g., private cars, buses, etc. The idea of “dedicated bus lane” is proposed and implemented in many cities worldwide to

\* Corresponding author at: Institute of Transport Studies, Department of Civil Engineering, Monash University, Australia.

E-mail address: [nan.zheng@monash.edu](mailto:nan.zheng@monash.edu) (N. Zheng).

promote public transport and divert car users to higher-occupied modes. The dedicated lanes give buses absolute priority without being interrupted by the traffic flow from other lanes. This is particularly helpful in the case of congestion where buses can pass the queues without experiencing significant delays. In practice, however, it may not be always possible to make one or more such lane(s) available everywhere. Also it is uneconomical to keep the same proportion of dedicated lanes in peak and non-peak periods (see Viegas and Lu, 2001 for a synthesis analysis). Therefore, the optimal design of dedicated bus lanes requires efficient distribution of the road capacity over both spatial and temporal horizons.

Attempts have been made to include different elements into the design of the dedicated lanes. For example, Basso et al. (2011) took into consideration mode choice, bus frequency, vehicle size, spacing optimization, and Tirachini and Hensher (2011) examined the effect of combining dedicated lanes with different fare-collecting schemes in the isolated bus system. While the factors that contributed to the delays of a bus trip are modelled in detail, neither of the models captures the multimodal traffic dynamics and allows for a time-varying road space distribution scheme. With the developments of advanced technologies and reductions in hardware costs, the real-time traffic signal controls give rise to the flexible dedicated lane design in different time periods during the day. Viegas and Lu (2001, 2004) and Eichler and Daganzo (2006) proposed and evaluated the intermittent bus lane signal scheme. But the interplays among dedicated lane, bus service operation, and traffic dynamics are not considered. In a similar manner, Guler and Menendez (2015), Guler et al. (2016) and He et al. (2016, 2017) introduced and analyzed pre-signals to provide a bypass of the car queues for buses in a mixed environment.

The works by Daganzo et al. (2012) and Gonzales and Daganzo (2012) provided an alternative approach for analyzing the relationship between land use and traffic congestion. These works linked the Macroscopic Fundamental Diagram (MFD) that describes the traffic dynamics with various features of transport infrastructures, and enabled the development of integrated infrastructure planning and transport operation strategies. The macroscopic traffic models are recently studied extensively, while the MFD with dynamic features is empirically observed by Geroliminis and Daganzo (2008), relating traffic flow (veh/hr) and traffic density (veh/km) at the level of an urban region. An interested reader can refer to, e.g., Mariotte et al. (2017), Saeedmanesh and Geroliminis (2017), Zhong et al. (2017), Loder et al. (2017), and Alsalhi et al. (2018) for reviews of recent developments in MFD. In contrast with other models of traffic congestion, e.g., the widely-used link performance function developed by the US Bureau of Public Roads (BPR, 1964) where congestion is assumed to depend only on the current demand without any memory for the level of congestion at previous times, the MFD captures the state-dependent hypercongestion where flow decreases as density increases beyond a critical value of density. Such treatment is found to be consistent with the physics of traffic congestion at network level, and it allows for the tractable modeling of intra-day traffic dynamics. As the MFD is jointly influenced by the topological characteristics of a network (other properties include traffic signals, multimodality etc., see in Daganzo and Geroliminis, 2008; Leclercq and Geroliminis, 2013; Saberi et al., 2014a, 2014b; Mariotte et al., 2017; Leclercq et al., 2017; Zockaei et al., 2018), congestion dynamics can be connected to the road space allocation between cars and public transport, and other operational characteristics. For example, under the framework of MFD, Tsekeris and Geroliminis (2013) studied the optimal selection of city size without considering multiple transportation modes. Zheng and Geroliminis (2013) analyzed the optimal space distribution of road space to different modes and its impact on congestion dynamics. The latter study shares common interests with this paper, but a challenging extension is to integrate bus operator's reaction strategies into the analysis.

Few studies in the literature have integrated the bus service optimization into the interaction among multiple travel modes with considering the properties of traffic flow. For example, Geroliminis et al. (2014) and Chiabaut et al. (2014) introduced the passenger macroscopic fundamental diagram (p-MFD) to analyze the flow interactions from different travel modes, but the characteristics of the bus operation, e.g., bus fares and service frequencies, are not assumed as exogenously given. Li et al. (2012) and Tirachini et al. (2014) studied the optimal bus frequency and fare when buses share the same road with cars, and Verbas et al. (2015, 2016) investigated the frequency setting in particular. However, none of these studies considered the effect of road capacity changes. In the context of road capacity expansion, Zhang et al. (2014, 2016) showed that transit operators with different objectives respond distinctly to the expansion of road capacity. They found that the transit service may deteriorate and users of both modes get worse-off after the road expansion. The implementation of dedicated bus lane has invasive impact on the road capacity shared by the bus mode. It is not necessarily true that the bus service will be more attractive; instead, the dedicated lane can lead to counter-productive result. This makes it intriguing to examine how the optimal space distribution should take into account the effect of bus service adjustments.

The optimal public transit operation and policy issues have been extensively studied in the single mode context. The seminal work by Mohring (1972) identified that the optimal bus frequency for welfare maximization is proportional to the ridership. This approach has been subsequently extended to the multimodal systems which jointly optimizes service frequency of public transport and other variables (e.g., bus size and fares in Basso and Jara-Díaz, 2012; road capacity, bus frequency and fare in Zhang et al., 2014, 2016). While most of the previous studies treat the bus running speed as an exogenous parameter, the congestion effect is often overlooked. In fact, the congestion induced by the frequent presence of buses and patronage slows down bus traffic and causes delays at bus stops when the demand is high. This aspect has been captured by a microeconomic model proposed by Ahn (2009), who discussed on the congestion interaction between the cars and buses when sharing the same road capacity. Tirachini and Hensher (2011) then developed a steady-state model to calculate bus delays through a detailed analysis of the formation of bus queues at stops; yet they assume the demand for bus is not influenced by the service quality. Recent works by Geroliminis et al. (2014) and Loder et al. (2017) developed the three-dimension Macroscopic Fundamental Diagram (3D-MFD) for representing the congestion dynamics in a bi-modal network where cars and buses share the same space. The space-mean flow or speed of vehicles (or passengers) is a function of both car and bus accumulation, with different degree of influence in the network performance (one additional bus has a larger effect in speed reduction compared to a car). The authors provided a tool for evaluating network performance when different buses serve in the network, from both vehicular and passenger point of view. Most of the aforementioned studies either regard mode choice as

exogenous or ignore the effect of land-use properties. While in practice, on one hand travelers can change their mode choices on the demand side given the service level of all the alternative modes, and on the other, the transport authority can modify the road space distribution to different modes on the supply side which influences the choice of the travelers. A systematic approach is required to reflect these interactions.

Motivated by the discussion above, this paper proposes a macroscopic framework for integrated modeling and optimization of the multimodal urban transport infrastructure capacity management and operations. The MFD-based system dynamic model is combined with the classic equilibrium-based model, aiming to present an approach that is not only consistent with the physics of traffic but also capable of identifying analytical characteristics of multimodal interactions. Under such framework, the dynamic user equilibrium with separated road space is formulated and the system-wide effects of the road space distribution is examined in a driving and bus two-mode system. The mode choice model and dynamic user equilibrium are integrated within the MFD dynamics, reflecting both vehicle and passenger dynamics of an urban area. In particular, the choice of bus depends decisively on the running speed of the buses which is affected by the bus lane capacity and the travel demand. The optimal bus fare and service frequency are firstly recognized for given space allocation in the steady state. Then we examine how the bus operator responds to space changes and how the system cost is influenced by the road space allocation. The model is then applied to a dynamic framework where the space distribution changes with the varying demand. Through numerical analysis, optimal bus operation strategies are investigated, the service adjustments at peak and off-peak periods are revealed, and the optimal space allocation strategy with responsive bus service is explored.

The reminder of this paper is organized as follows. [Section 2](#) formulates the dynamic user equilibrium with MFD representation. [Section 3](#) analyzes the optimal bus service for given road space separation in the steady state. [Section 4](#) examines the system-wide effects of road space allocation with responsive bus service. [Section 5](#) explores the optimal bus service and space allocation in the dynamic framework via a comprehensive numerical case study. Conclusions are given in the last section.

## 2. Model framework

Consider a downtown area where the total road capacity is divided for private cars and buses. The total length of available travel lane in the network is denoted by  $L(\text{km})$ . Let  $\lambda$  denote the fraction of the total length assigned for dedicated bus lanes, and  $0 < \lambda < 1$ . We assume that  $\lambda$  is continuously adjustable and thus the ensuing analysis is relevant in the whole range of  $\lambda$ . Due to practical feasibility, the change of space allocation should not appear too frequently on the urban road. We consider it is indeed manageable to give one or more lanes for buses in a fraction of roads in the network during the peak period. In other words, some lanes are designated as bus only during the peak period, while after the rush hours, other vehicles can use these lanes as well.

Denote  $D(t)$  the total travel demand (the number of travelers departing for a trip) at time  $t$ . We assume every trip has the identical length  $l$ , which can be estimated with historical data. This distance of “trip length” can be a part of the whole trip (cross-region trips) or an entire trip that originates and ends in the targeted region. Empirical analysis (see in [Geroliminis and Daganzo, 2008](#)) revealed that the value appeals to be stable with moderate fluctuations.

Travelers choose between two alternative travel modes, i.e., car and bus (represented by subscripts ‘a’ and ‘b’ respectively), based on their generalized travel costs. We will elaborate the formulation of the travel cost in the later text. Let  $x_a(t)$  and  $x_b(t)$  respectively represent the demand for car and bus (the resulting modal split) at time  $t$ , where

$$x_a(t) + x_b(t) = D(t). \quad (1)$$

### 2.1. Driving mode

The congestion of the car traffic in the concerned area is considered homogeneously distributed over space and exhibits an MFD with low scatter. Note that it is not difficult to extend our formulation to networks with congestion heterogeneity, e.g. by partitioning such network into regions of the same congestion patterns (see [Saeedmanesh and Geroliminis, 2016](#)). Denote  $n_a$  the car accumulation (number of cars in the system),  $k_a$  the car traffic density (number of vehicles in unit length of road), and then

$$n_a = (1-\lambda)L \cdot k_a. \quad (2)$$

Since the average running speed  $v_a$  of all the vehicles in the area decreases with the traffic density  $k_a$ , it hence decreases with the accumulation  $n_a$  and increases with the capacity allocated for cars, i.e.,

$$v_a = v_a(n_a; \lambda), \quad \partial v_a / \partial n_a < 0, \quad \partial v_a / \partial \lambda < 0 \quad (3)$$

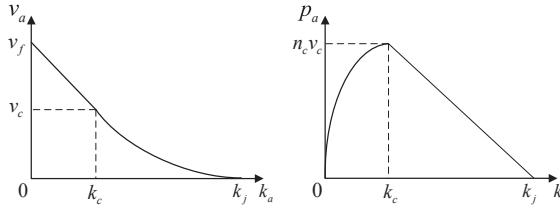
The production (vehicle kilometers traveled per unit time) of the system is

$$p_a(n_a; \lambda) = n_a v_a(n_a; \lambda), \quad (4)$$

and the outflow (rate at which vehicles reach their destination) of the system is

$$O_a(n_a; \lambda) = \frac{p_a(n_a; \lambda)}{l} = \frac{n_a v_a(n_a; \lambda)}{l}. \quad (5)$$

At the network level, the dynamics of the car traffic are captured by the differential equation of the mass conservation:



**Fig. 1.** The traffic flow model for car traffic ( $n_c = (1-\lambda)L \cdot k_c$ ,  $v_f = 2v_c$ ).

$$\frac{dn_a(t)}{dt} = x_a(t) - O_a(t), \quad n_a(0) = n_{a0} \quad (6)$$

Considering the complexity of the problem, which is established in the literature, we introduce the linear/hyperbolic MFD (as shown in Fig. 1) to characterize the dynamics of car traffic. The benefit of such a simplified MFD model is twofold. As it appears, either the speed or the production is in the linear form hence it can ease the algebra and maintain the analytical tractability of the model. More widely, this model can nicely reflect the general monotonicity properties of MFDs identified from empirical evidence (e.g., Geroliminis and Daganzo, 2008). This implies that the major results of subsequent analysis (except the absolute solutions), related to the monotonicity properties of MFDs, can be generalized to other settings. Similar simplified forms of MFDs have been used in other analytical studies, such as Daganzo (2007), Haddad et al. (2013), Arnott (2013), Liu and Geroliminis (2016), Yang et al. (2017), and Lamott and Geroliminis (2017).

Mathematically, the speed-density relationship is given by:

$$v_a(k_a) = \begin{cases} -\frac{v_c}{k_c}k_a + 2v_c, & \text{where } 0 < k_a \leq k_c \\ \frac{k_c v_c}{k_j - k_c} \left( \frac{k_j}{k_a} - 1 \right), & \text{where } k_c < k_a \leq k_j \end{cases}. \quad (7)$$

where  $k_c$ ,  $k_j$ , and  $v_c$  represent the critical density, jam density, and critical speed respectively. The critical parameters offer important information for traffic management, as a network becomes oversaturated and hypercongestion forms when the critical values are exceeded. Specifically, where traffic density  $k_a$  is smaller than the critical density  $k_c$ , the traffic speed  $v_a$  is greater than the critical speed  $v_c$ , and the traffic condition here is referred to as the ‘uncongested regime’. Where  $k_a$  exceeds  $k_c$ ,  $v_a$  falls below  $v_c$ , and it is referred to as the ‘hypercongested regime’. When  $k_a$  approaches the jam density  $k_j$ ,  $v_a$  approaches zero, i.e., the ‘gridlock’ occurs. The speed-density function defined in Eq. (7) is continuous but non-differentiable at the critical density. The traffic speed  $v_a$  monotonically decreases with the density  $k_a$ , with a kink at  $k_a = k_c$ .

Substituting Eqs. (2) and (7) into (4) and with some manipulations, we have the following traffic production as a function of traffic density:

$$p_a(k_a) = \begin{cases} -\frac{(1-\lambda)Lv_c}{k_c}(k_a - k_c)^2 + n_c v_c, & \text{where } 0 < k_a \leq k_c \\ \frac{n_c v_c}{k_j - k_c}(k_j - k_a), & \text{where } k_c < k_a \leq k_j \end{cases} \quad (8)$$

where  $n_c = (1-\lambda)L \cdot k_c$ , representing the car accumulation at the critical density. Eq. (8) prescribes that in the uncongested regime, the traffic production  $p_a$  quadratically increases with  $k_a$ , and reaches the maximum  $p_a^{\max} = n_c v_c$  where  $k_a = k_c$ . The maximum coincides with that of the hypercongested regime where  $k_a$  approaches  $k_c$ . When  $k_a$  exceeds  $k_c$ ,  $p_a$  decreases with  $k_a$  and approaches zero when  $k_a$  reaches jam density. The production-density function, similar to the speed-density function, is thus continuous but non-differentiable at  $k_a = k_c$ .

The generalized travel cost of the car user who enters the system at time  $t$  is:

$$c_a(t; \lambda) = \tau_a + \beta \cdot tt_a(n_a(t); \lambda), \quad (9)$$

where  $\tau_a$  is the monetary cost of the car user,  $\beta$  is the value of in-vehicle travel time, and  $tt_a$  is the travel time by car.  $tt_a$  is calculated by trip length over speed, i.e.,  $tt_a(n_a; \lambda) = l/v_a(n_a; \lambda)$ , where  $n_a(t)$  is the instantaneous car accumulation at time  $t$ . While dynamic conditions might influence the accuracy of instantaneous estimation of travel time, such an approach is considered for its analytical tractability (errors are not expected to be more than 10%, see Yildirimoglu and Geroliminis, 2013 or Liu and Geroliminis, 2016).

## 2.2. Bus mode

With dedicated bus lanes, we assume that the bus traffic is independent of the car traffic. The average speed of buses is captured by:

$$v_b(t; \lambda) = v_{b0} - \delta(\lambda)x_b(t). \quad (10)$$

The first term on the RHS of Eq. (10) is the free-flow speed of buses and is assumed as a constant,  $v_{b0}$ . The second term is the product of the instantaneous demand  $x_b(t)$  and the factor  $\delta(\lambda)$  determined by the road space allocation. Eq. (10) captures the

relationship among bus speed, demand for bus, and road space allocation.<sup>1</sup> For a given space allocation  $\lambda$ , Eq. (10) is a linear approximation of the relationship between bus speed and travel demand. The bus speed decreases when more people choose the bus mode, because longer dwelling time is needed at each bus stop for boarding and alighting when the ridership is higher.

For a given demand  $x_b$ , more space allocated for bus mode means that the demand can be more spread spatially and that a particular bus can be less delayed by boarding and alighting at each bus stop. Thus, the factor  $\delta(\lambda)$  represents the marginal effect of travel demand on bus speed depending on the space allocation. With more space allocated for buses, the marginal effect diminishes, so that  $\delta$  decreases with  $\lambda$ , i.e.,  $\delta' = d\delta/d\lambda < 0$ . To ensure a strictly positive bus running speed for any demand, we assume a positive bus running speed at the minimum of  $\lambda$  and the maximum of  $D$ , i.e.,  $v_{b0} - \delta(\lambda)D \geq 0$ .

We measure the generalized travel cost of a bus rider who enters the system at time  $t$  by the sum of bus fare, in-vehicle travel time cost, and waiting time cost at bus stops:

$$c_b(t) = \tau_b(t) + \beta \cdot t_{t_b}(t) + \alpha \cdot w(f), \quad (11)$$

where  $\tau_b$  is the bus fare,  $t_{t_b}$  is the in-vehicle travel time by bus,  $\beta$  is the value of in-vehicle travel time,  $w(f)$  is the waiting time, and  $\alpha$  is the value of waiting time. Given the trip length  $l$  and average speed  $v_b$ , the in-vehicle travel time by bus is  $t_{t_b}(t) = l/v_b(t)$ . The average waiting time for a bus is approximated by the half of the headway, which is an accurate estimation if buses run at schedule (by definition headway is the inverse of frequency)  $w(f) = 1/2f$ . Both the bus fare and the service frequency are decision variables for the bus operator, and they are both adjustable in response to the change in road space allocation. Throughout the analysis of this paper, it is assumed that there is no bus bunching and that the combination of bus vehicle size and service frequency is large enough to carry all the travelers waiting on the bus stop, following many other analytical studies aiming at high-level multimodal problems, e.g., Ahn (2009), Tirachini and Hensher (2011), and Zhang et al. (2014, 2016).<sup>2</sup>

It is noteworthy that an implicit relationship may exist between the space allocation and the bus service accessibility. As more space allocated to dedicated bus lanes, a denser bus service network can be developed with more bus stops and in broader areas. Such development leads to different operations of the transit system which entails detailed engineering design at lower-level (e.g. bus stop redesign). Compared to the adjustments of service frequency and ticket price which are more flexible, this treatment is not trivial and it may cause issues that lose the advantage of the macroscopic approach. Given that our intention in this paper is to focus on the higher-level problem and that the current model is already highly complex, we treat the access cost as exogenous and thus omit this term in Eq. (11). In future studies, the quantification of this term can be carried out based on empirical data and feeds into the modeling framework.

### 2.3. Dynamic user equilibrium

The above described traffic dynamics of both modes rely on the number of vehicles/passengers entering each system, which is in turn driven by the instantaneous modal split. At any given time point  $t$ , a total of  $D(t)$  trips generate in the system and split into two groups that feed in the two modes as instantaneous demand/inflow, denoted by  $x_a(t)$  and  $x_b(t)$  respectively. Travelers make mode choice based on the generalized travel costs of the two modes, which are both time-dependent and flow-dependent.

By assuming deterministic mode choice, the equilibrium modal split at time  $t$  is pinned down by

$$[c_i(t) - c_{\min}(t)] \cdot x_i(t) = 0, \quad i = a, b \quad (12)$$

where  $c_{\min}(t) = \min\{c_a(t), c_b(t)\}$ , representing the minimum travel cost between alternatives. The equilibrium condition Eq. (12) includes both interior and boundary equilibria. Where both modes are used, the interior user equilibrium is reached when the generalized travel costs of the used modes are equal, i.e.,  $c_a(t) = c_b(t)$ . Where the boundary equilibrium prevails, no one chooses the mode with the larger travel cost, i.e.,  $x_i(t) = 0$  if  $c_i(t) > c_j(t)$ , with  $i, j = a, b, i \neq j$ . The above condition serves as the basis for the subsequent equilibrium analysis in either static or dynamic context.<sup>3</sup>

### 2.4. System-wide welfare and cost measurements

We now defined the welfare measurements concerned in this paper. From the bus operator's perspective, the profit (TBP) is part of the fare box revenue that outweighs the operational expenses:

$$TBP = \int_0^T [\tau_b(t)x_b(t) - K(f(t))] dt, \quad (13)$$

where  $K(f(t))$  represents the operational cost associated with the service frequency. We assume the constant marginal operational cost for each bus movement, and the bus operator's capital and operation expenses can be captured by  $K(f) = K_0 + K_1 f$ , in which  $K_0$  denotes the capital cost (presumably fixed in the short run) and  $K_1$  characterizes the marginal operational cost (associated with each

<sup>1</sup> To simplify the algebra, the headway of buses is not involved in this formulation. However, the headway effect can be partially captured by the endogenous relationship between the service frequency and the demand.

<sup>2</sup> Where boarding feasibility is an issue, it can be readily incorporated into the current framework by adding a demand-driven disutility term in the bus generalized travel cost to reflect the boarding impedance and/or on-board comfort in Zheng and Geroliminis (2013) and de Palma et al. (2017).

<sup>3</sup> The proposed bimodal equilibrium framework does not include the total demand elasticity towards the equilibrium cost, namely departure time shifts and the induced travels are not incorporated. Recent research works such as Arnott and Yan (2000), Zhang et al. (2016), Liu and Geroliminis (2016), and Lamott and Geroliminis (2017) shed lights on modelling and integrating these factors.

movement). Similar linear operational cost function has been adopted in related literature such as Ahn (2009) and Basso and Jara-Díaz (2012). Since the service frequency is adjustable over time, the operational cost function  $K(f)$  is also time-dependent.

With regard to users, the total travel cost (TTC) of all the users can be calculated by

$$TTC = \int_0^T [x_a(t)c_a(t) + x_b(t)c_b(t)]dt. \quad (14)$$

From the system's point of view, the total system cost (deadweight loss, TSC) is the sum of operators' and users' losses:

$$TSC = TTC - TBP = \int_0^T [x_a(t)c_a(t) + x_b(t)(c_b(t) - \tau_b(t)) + K(f(t))]dt, \quad (15)$$

where the construction and maintenance costs associated with the change in road space distribution are excluded from consideration.

Given homogeneous travel preference, the total system cost (TSC) is an equal but opposite measure of the net economic benefit (NEB, sum of producer's surplus and consumers' surplus) for the system, and maximizing NEB is equivalent to minimizing TSC (Varian, 1992).

### 3. Properties of the stationary equilibrium

Utilizing the system described in the last section, we now look into properties of the static equilibrium where the bimodal system reaches the steady state. We focus on the equilibrium individual travel cost and modal split properties for given road space allocation and bus service parameters.

In terms of car traffic, the steady state is reached when the marginal change in traffic accumulation is constant, i.e.,  $dn_a(t)/dt = 0$ , where inflow equals outflow  $x_a(t) = O_a(t)$ . According to Eq. (6) we have the following relationship among inflow (demand)  $x_a$ , traffic accumulation  $n_a$ , and traffic speed  $v_a$ :

$$x_a(t) = \frac{n_a(t)v_a(t)}{l}. \quad (16)$$

With the steady state condition Eq. (16), we rewrite the car traffic speed defined in Eq. (7) in terms of demand  $x_a$  in the two regimes of the MFD:

$$v_a = \begin{cases} \frac{\sqrt{n_c v_c (n_c v_c - l x_a)}}{n_c} + v_c, & \text{where } 0 < n_a \leq n_c \text{ (or } 0 < k_a \leq k_c\text{)} \\ \frac{l x_a}{n_j - \frac{n_j - n_c}{n_c v_c} l x_a}, & \text{where } n_c < n_a \leq n_j \text{ (or } k_c < k_a \leq k_j\text{)} \end{cases} \quad (17)$$

and the traffic accumulation  $n_a$  in terms of  $x_a$ :

$$n_a = \begin{cases} n_c - \frac{\sqrt{n_c v_c (n_c v_c - l x_a)}}{v_c}, & \text{where } 0 < n_a \leq n_c \text{ (or } 0 < k_a \leq k_c\text{)} \\ n_j - \frac{n_j - n_c}{n_c v_c} l x_a, & \text{where } n_c < n_a \leq n_j \text{ (or } k_c < k_a \leq k_j\text{)} \end{cases} \quad (18)$$

for which an explicit derivation is provided in Appendix A. Note that the steady state of traffic accumulation in the proposed model is state-dependent, with respect to the critical density, whereas most existing works yields a single analytical solution. Nevertheless, the derivation of Eq. (18) gets inspiration from several studies in this direction such as the one by Daganzo et al. (2012).

In Eqs. (17) and (18),  $v_a$  and  $n_a$  are given by implicit functions defined in each regime/branch of the MFD – when the underlying car traffic is in the uncongested regime ( $0 < n_a < n_c$  or equivalently  $0 < k_a < k_c$ ) the first formula prevails, and the other one otherwise. For simplicity, we refer to the MFD regime by the range of  $n_a$  in the subsequent analysis. Based on Eqs. (17) and (18), we examine the marginal effects of  $x_a$  on  $n_a$  and  $v_a$  and obtain the following lemma:

**Lemma 1..** *When the concerning bimodal system reaches the steady state,*

- (i)  $\frac{\partial n_a}{\partial x_a} > 0$  and  $\frac{\partial v_a}{\partial x_a} < 0$ , where  $0 < n_a \leq n_c$ ;
- (ii)  $\frac{\partial n_a}{\partial x_a} < 0$  and  $\frac{\partial v_a}{\partial x_a} > 0$ , where  $n_c < n_a \leq n_j$ .

The proof of Lemma 1 is straightforward from Eqs. (17) and (18) and is thus omitted. This lemma reflects the steady-state relationship between inflow  $x_a$  and car accumulation  $n_a$  (or speed  $v_a$ ), and signifies the comparison between any two equilibrium states. Where the traffic is in the uncongested regime ( $0 < n_a \leq n_c$ ), the equilibrium state with larger car inflow has the larger accumulation and lower speed. With regard to two equilibrium states in the hypercongested regime ( $n_c < n_a \leq n_j$ ), the one with larger inflow has the lower accumulation and higher speed. The effects are distinct in different traffic conditions. However, it is noteworthy that the effects are in line with the relationship between outflow  $O_a$  and  $n_a$  (or  $v_a$ ), as the steady-state condition primarily requires outflow  $O_a$  equals inflow  $x_a$ . Daganzo (2007) established the steady-state (equilibrium) behavior of traffic outflow – outflow increases with accumulation in the uncongested regime and decreases in hypercongestion. Therefore, the steady-state relationship described in Lemma 1 is consistent with the general properties of traffic flow models. As the advancement from the results of Daganzo (2007), Lemma 1 focuses on the inflow rate (rather than exit rate) which is more directly linked with the modal split, and includes the effect of flow not only on accumulation but also on speed.

It is noteworthy that the speed-density function defined in Eq. (7) is not differentiable at the critical density  $k_c$ . Thus, none of functions  $p_a(k_a)$ ,  $n_a(x_a)$ , or  $v_a(x_a)$  is differentiable at  $k_c$  or  $n_c$ . The partial derivatives of  $n_a$  and  $v_a$  with respect to  $x_a$  are defined in line with the left limit of functions  $n_a(x_a)$  and  $v_a(x_a)$  when  $n_a$  approaches  $n_c^-$ . The resulting  $\partial n_a/\partial x_a$  and  $\partial v_a/\partial x_a$  are discontinuous at the boundary of the two regimes of the MFD. This property prevails in many aspects of stationary equilibrium analysis. Since the boundary traffic condition between uncongested and hypercongested regimes (or how does traffic condition transmit from one regime to another) of the MFD is primarily debatable and is out of the scope of this paper, the properties of the stationary equilibrium are examined respectively for the two regimes.

To investigate the properties of an interior bimodal equilibrium in the steady state, we assume both modes are utilized and the equilibrium is reached where

$$c_a = c_b \Leftrightarrow \tau_a + \beta \cdot \frac{l}{v_a} = \tau_b + \beta \cdot \frac{l}{v_b} + \alpha \cdot \frac{1}{2f}. \quad (19)$$

The underlying assumptions for the interior equilibrium are specified in the Appendix B. With  $D$  denoting the total demand for the two modes in a time unit in the steady state, the equilibrium mode share of car is determined by the solution of equation  $g(x_a) = c_b - c_a = 0$ , and

$$g(x_a) = \begin{cases} \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{\beta l}{v_{b0} - \delta(D - x_a)} - \frac{\beta ln_c}{\sqrt{n_c v_c (n_c v_c - lx_a)} + n_c v_c}, & \text{where } 0 < n_a \leq n_c \\ \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{\beta l}{v_{b0} - \delta(D - x_a)} - \beta \frac{n_j - n_c}{n_c v_c} lx_a, & \text{where } n_c < n_a \leq n_j \end{cases} \quad (20)$$

For given  $\lambda$ ,  $f$ ,  $\tau_b$ , and  $D$ , the equilibrium mode split  $x_a$  and  $x_b$  can be uniquely determined in either the uncongested regime ( $0 < n_a \leq n_c$ ) or the hypercongested regime ( $n_c < n_a \leq n_j$ ). The proof is provided in the Appendix B. As such, the car usage  $x_a$  can be regarded as a function of  $f$ ,  $\tau_b$ , and  $\lambda$ , i.e.,  $x_a = x_a(f, \tau_b, \lambda)$ . Taking partial derivatives of the two sides of Eq. (19) with respect to  $f$ ,  $\tau_b$ , and  $\lambda$ , and with some manipulations, we obtain the marginal effects of  $f$ ,  $\tau_b$ , and  $\lambda$  on  $x_a$  respectively (an explicit derivation is provided in Appendix C):

$$\frac{\partial x_a}{\partial f} = -\frac{\alpha}{2\beta\gamma f^2}, \quad (21)$$

$$\frac{\partial x_a}{\partial \tau_b} = \frac{1}{\beta\gamma}, \quad (22)$$

$$\frac{\partial x_a}{\partial \lambda} = \frac{\theta}{\gamma}, \quad (23)$$

where

$$\theta = \frac{1}{v_a^2} \frac{\partial v_a}{\partial \lambda} - \frac{1}{v_b^2} \frac{\partial v_b}{\partial \lambda}, \quad \gamma = \frac{1}{v_b^2} \frac{\partial v_b}{\partial x_a} - \frac{1}{v_a^2} \frac{\partial v_a}{\partial x_a}$$

**Proposition 1..** In steady state, the marginal effects of  $f$ ,  $\tau_b$ , and  $\lambda$  on the equilibrium car usage satisfy

$$\frac{\partial x_a}{\partial f} < 0, \quad \frac{\partial x_a}{\partial \tau_b} > 0, \quad \text{and} \quad \frac{\partial x_a}{\partial \lambda} < 0 \quad (24)$$

if and only if Condition (25) or (26) is valid:

$$0 < n_a \leq n_c \quad (25)$$

$$n_c < n_a \leq n_j, \quad \delta > \frac{\partial v_a}{\partial x_a} \frac{v_b^2}{v_a^2}, \quad \text{and,} \quad \delta' < -\frac{\partial v_a}{\partial x_a} \frac{x_a}{(1-\lambda)x_b} \frac{v_b^2}{v_a^2} \quad (26)$$

**Proof..** Clearly,  $\frac{\partial v_b}{\partial x_a} = \delta > 0$  and  $\frac{\partial v_b}{\partial \lambda} = -x_b \delta' > 0$ , where  $\delta$  is defined as a decreasing function of  $\lambda$ , and  $\delta' = d\delta/d\lambda < 0$ . We thus have

$$\frac{\partial v_a}{\partial \lambda} = \begin{cases} -\frac{v_c lx_a}{2(1-\lambda)\sqrt{n_c v_c (n_c v_c - lx_a)}} = -\frac{v_c lx_a}{2n_c(1-\lambda)(v_a - v_c)} < 0, & \text{if } 0 < n_a \leq n_c \\ \frac{Lk_j lx_a}{\left(n_j - \frac{n_j - n_c}{n_c v_c} lx_a\right)^2} = \frac{Lk_j v_a^2}{lx_a} > 0, & \text{if } n_c < n_a \leq n_j \end{cases} \quad (27)$$

When  $0 < n_a \leq n_c$ ,  $\frac{\partial v_a}{\partial x_a} < 0$ ,  $\gamma = \frac{1}{v_b^2} \frac{\partial v_b}{\partial x_a} - \frac{1}{v_a^2} \frac{\partial v_a}{\partial x_a} > 0$ , and  $\theta = \frac{1}{v_a^2} \frac{\partial v_a}{\partial \lambda} - \frac{1}{v_b^2} \frac{\partial v_b}{\partial \lambda} < 0$ . Substituting into Eqs. (21)(22)(23), we have  $\frac{\partial x_a}{\partial f} < 0$ ,  $\frac{\partial x_a}{\partial \tau_b} > 0$ , and  $\frac{\partial x_a}{\partial \lambda} < 0$ .

When  $n_c < n_a \leq n_j$ ,  $\frac{\partial v_a}{\partial x_a} = \frac{n_j v_a^2}{lx_a^2}$ , and then

$$\frac{\partial x_a}{\partial f} < 0 \quad \text{or} \quad \frac{\partial x_a}{\partial \tau_b} > 0$$

$$\Leftrightarrow \delta > \frac{n_j v_b^2}{l x_a^2}. \quad (28)$$

And

$$\begin{aligned} \frac{\partial x_a}{\partial \lambda} &< 0 \\ \Leftrightarrow \left( \delta - \frac{n_j v_b^2}{l x_a^2} \right) \left( \delta' + \frac{L k_j v_b^2}{l x_a x_b} \right) &< 0. \end{aligned} \quad (29)$$

Eqs. (28) and (29) are simultaneously valid if and only if

$$\delta > \frac{\partial v_a}{\partial x_a} \frac{v_b^2}{v_a^2}, \text{ and, } \delta' < -\frac{\partial v_a}{\partial x_a} \frac{x_a}{(1-\lambda)x_b} \frac{v_b^2}{v_a^2}. \square$$

Condition (25) in Proposition 1 shows that when the car traffic is in the uncongested regime, the mode share of car increases if the service level of the alternative decreases, e.g. service frequency drops, bus fare goes up, or bus lane shrinks. Similar result applies to the case where car travel time is characterized by a general function as in Zhang et al. (2014). However, since the congestion effect in the bus mode is not captured, Zhang et al. (2014) can be viewed as a special case of this paper where the car travel time is a function increases with traffic volume and decreases with road capacity (consistent with the uncongested regime of the MFD).

When the car traffic condition is on the right branch of the MFD (hypercongested regime), higher inflow/outflow corresponds to higher speed of car traffic (shown in Lemma 1 (ii)). While on the bus side, the traffic speed decreases with the mode share of bus. As the consequence, increasing bus frequency or reducing fare can increase bus ridership only if the marginal effect of modal split on bus speed outweighs its effect on car speed. The effect of the road capacity allocation, both  $v_a$  and  $v_b$  increase with parameter  $\lambda$  (refer to the Proof of Proposition 1). It follows that only if the marginal effect of the space fraction is large enough on bus speed to offset the effect on car traffic, enlarging bus lane space can improve bus ridership. Zhang et al. (2016) proved that transit ridership and service quality may not move in tandem as in Mohring's original model (in Mohring, 1972) by considering the effects of road expansion and imperfect substitutes. Condition (26) in Proposition 1 shows that neither improving bus service nor providing more space for buses can necessarily increase the bus ridership even if the two modes are perfect substitutes.

#### 4. The profit-maximizing bus service with road space separation

Section 3 establishes the effects of bus service and road space allocation on traffic condition and equilibrium modal split. The frequency and fare are both decision variables of the bus operator and are both adjustable in response to the road space allocation. The operator's decision depends on its goal. This section considers the bus operator's decision aiming for profit-maximization, treating the space allocation as exogeneous. We investigate the optimal bus frequency and fare under a given space allocation  $\lambda$ . According to the definition in Eq. (13), the profit of the bus operator (BP) within a unit time in the steady state is

$$BP = \tau_b x_b - K(f). \quad (30)$$

First-order conditions for the optimal frequency and fare ( $f^*$ ,  $\tau_b^*$ ) are:

$$f^*: \tau_b \frac{\partial x_b}{\partial f} = K', \quad (31)$$

$$\tau_b^*: \tau_b \frac{\partial x_b}{\partial \tau_b} + x_b = 0. \quad (32)$$

Eq. (31) prescribes that the marginal revenue from increasing frequency balances the marginal increase in cost. Eq. (32) dictates that the marginal revenue from adjusting the fare is zero. The Hessian matrix of the optimization problem is negative semi-definite under Condition (57) derived in Appendix D. Denote  $f^u$  and  $\tau_b^u$  the optimal frequency and fare when the car traffic is in the uncongested regime, and  $f^h$  and  $\tau_b^h$  the optimal frequency and fare when the car traffic is in hypercongestion. We then have the following proposition:

**Proposition 2..** Under a specific road space allocation  $(\lambda, 1-\lambda)$ , the profit-maximizing bus frequency and fare satisfy:

- (i) if  $0 < n_a \leq n_c$ ,  $f^u = \sqrt{\frac{\alpha x_b}{2K'}}$ ,  $\tau_b^u = \beta l x_b \left( \frac{\delta}{v_b^2} + \frac{l v_c}{2(v_a - v_c) n_c v_a^2} \right)$ ;
- (ii) if  $n_c < n_a \leq n_j$ ,  $f^h = \sqrt{\frac{\alpha x_b}{2K'}}$ ,  $\tau_b^h = \beta l x_b \left( \frac{\delta}{v_b^2} - \frac{n_j}{l x_a^2} \right)$ ;
- (iii) for the same  $x_b$  and  $v_b$ ,  $f^u = f^h$  and  $\tau_b^u > \tau_b^h$ .

**Proof..** The results in (i) and (ii) can be derived by substituting Eqs. (21) and (22) into Eqs. (31) and (32), and (iii) follows from (i) and (ii).  $\square$

Note that these relations do not provide explicit expressions for  $f$  and  $\tau_b$  since the modal split depends on the service frequency

and fare. Nevertheless, these intermediate results have the following implications. Firstly, it is shown that the profit-maximizing bus service frequency replicates the “square root principle”. The “square root principle” in the classic public transport literature (e.g., Mohring, 1972; Jansson, 1984) dictates that the first-best service frequency is proportional to the square root of the ridership for social welfare-maximization. It leads to scale economies of public transport service when both users’ and operator’s costs are considered. The optimization of bus service considered in Proposition 2 does not directly include the users’ costs into the objective, however, the operator’s profit is driven by the modal split which implicitly relies on the users’ cost. Therefore, Proposition 2 shows that when mode choice is available the profit-maximizing bus service also exhibits increasing returns to scale.

Proposition 2(iii) compares the profit-maximizing service frequency and fare for the two regimes of the MFD of the car traffic. The speeds of cars are not comparable in different regimes of the MFD ( $v_a \geq v_c$  in uncongested regime and  $v_a < v_c$  in hypercongestion). Nevertheless, both bus running speed  $v_b$  and ridership  $x_b$  can be equal in the formulas for  $f^u$  and  $f^h$ , or  $\tau_b^u$  and  $\tau_b^h$ . When this happens, it yields identical service frequency for the two regimes  $f^u = f^h$  and higher fare for the uncongested regime  $\tau_b^u > \tau_b^h$ . This is because the optimal frequency only depends on ridership, and the optimal fare is where the marginal revenue from adjusting the fare is zero,  $\tau_b = |x_b| / \left| \frac{\partial x_b}{\partial \tau_b} \right|$ . According to Eq. (22), the loss in ridership is less in the uncongested regime when the fare is marginally increased, i.e.,  $\left| \frac{\partial x_b}{\partial \tau_b} \right|_u < \left| \frac{\partial x_b}{\partial \tau_b} \right|_h$ . Therefore, the profit margin is larger in the uncongested regime than in hypercongestion.

## 5. The system-wide effect of road space allocation and Pareto-improving situation

Section 4 establishes the profit-maximizing bus fare and service frequency for given road space allocation. In practice the allocation of road space is usually decided by the transport authority at the higher level. There typically exists conflict in the objectives of public authority and private corporation. To unveil the interaction between the transport authority and bus operator in the context of road space separation, we consider a transport authority concerned with the total social cost controls the space distribution, whereas a profit-driven bus service operator determines the service frequency and ticket price. We examine the effect of road space allocation on bus operator’s profit and how system cost changes with the allocation. The Pareto-improving condition is then explored with which expanding the bus lane results in bus profit increase and system cost reduction at the same time (the win-win situation).

**Proposition 3..** Under the bus service frequency and fare established in Proposition 2, the bus operator’s profit BP increases with the proportion of road space distributed to bus lanes  $\lambda$  under Condition (25) or (26), otherwise BP decreases with  $\lambda$ .

**Proof..** The effect of a marginal change in space fraction can be evaluated by taking the total derivative of the profit function in Eq. (30) with respect to  $\lambda$ :

$$\begin{aligned} \frac{dBP}{d\lambda} &= -\tau_b \frac{dx_a}{d\lambda} + x_b \frac{d\tau_b}{d\lambda} - K' \frac{df}{d\lambda} \\ &= -\tau_b \left( \frac{\partial x_a}{\partial \lambda} + \frac{\partial x_a}{\partial f} \frac{df}{d\lambda} + \frac{\partial x_a}{\partial \tau_b} \frac{d\tau_b}{d\lambda} \right) + x_b \frac{d\tau_b}{d\lambda} - K' \frac{df}{d\lambda}. \end{aligned} \quad (33)$$

From Eqs. (31) and (32), we have

$$x_b = \tau_b \frac{\partial x_a}{\partial \tau_b}, \quad (34)$$

and

$$K' = -\tau_b \frac{\partial x_a}{\partial f}. \quad (35)$$

Substituting Eqs. (34) and (35) into Eq. (33) yields

$$\frac{dBP}{d\lambda} = -\tau_b \left( \frac{\partial x_a}{\partial \lambda} + \frac{\partial x_a}{\partial f} \frac{df}{d\lambda} + \frac{\partial x_a}{\partial \tau_b} \frac{d\tau_b}{d\lambda} \right) + \tau_b \frac{\partial x_a}{\partial \tau_b} \frac{d\tau_b}{d\lambda} + \tau_b \frac{\partial x_a}{\partial f} \frac{df}{d\lambda} = -\tau_b \frac{\partial x_a}{\partial \lambda} = \tau_b \frac{\partial x_b}{\partial \lambda}. \quad (36)$$

Therefore, BP increases with  $\lambda$  when  $\partial x_a / \partial \lambda$  is negative, i.e., under Condition (25) or (26).  $\square$

Proposition 3 shows that the marginal profit of bus operation associated with marginal bus lane capacity is proportional to the bus fare  $\tau_b$ , and its change direction is consistent with the marginal bus ridership  $\partial x_b / \partial \lambda$ .

We now look into the system cost. According to the definition given in Eq. (15), the system cost (SC) within a unit time is

$$SC = x_a c_a + x_b c_b - [x_b \tau_b - K(f)] = D c_a - x_b \tau_b + K(f). \quad (37)$$

Accordingly, the marginal effect of road space allocation on the system cost can be calculated by

$$\begin{aligned} \frac{dSC}{d\lambda} &= D \frac{dc_a}{d\lambda} - \frac{dBP}{d\lambda} \\ &= D \frac{\partial c_a}{\partial v_a} \left( \frac{\partial v_a}{\partial \lambda} + \frac{\partial v_a}{\partial x_a} \frac{dx_a}{d\lambda} \right) - \frac{dBP}{d\lambda} \end{aligned} \quad (38)$$

$$= D\beta l \left( \gamma - \frac{\delta}{v_b^2} \right) \frac{dx_a}{d\lambda} + D\beta l \left( \frac{\delta' x_b}{v_b^2} - \theta \right) + \frac{\theta \tau_b}{\gamma}, \quad (39)$$

Eq. (38) follows from Eq. (36). The first term of Eq. (38) captures the change of total travel cost of all travelers combining direct and indirect impacts. The change of space allocation directly leads to change in the lane kilometers available for cars – hence influencing the properties of the MFD and the traffic speed, captured by  $\partial v_a / \partial \lambda$ . At the same time, the change of space allocation reshapes the equilibrium modal split and indirectly influences the speed. The second term of Eq. (38) represents the change in the bus operator's profit.

Eq. (39) is obtained by substituting Eqs. (23) and (36) into Eq. (38). Eq. (39) implies that the change of SC is driven by the change of equilibrium mode share, or  $dx_a/d\lambda$ , which is in turn governed by the bus operator's response to the space change.

Zhang et al. (2014, 2016) have shown how the guideway transit operator reacts to the change of the road capacity. In an analogous manner, we now investigate the adjustment when the bus operator faces the change in road capacity distribution. Denote  $df/d\lambda$  and  $d\tau_b/d\lambda$  the adjusting rate of frequency and fare with respect to the space fraction respectively. Given Eqs. (31), (32) and (57), we have

$$\frac{dx_a}{d\lambda} + \frac{4K'f}{\alpha} \frac{df}{d\lambda} = 0, \quad (40)$$

$$\frac{d\tau_b}{d\lambda} - \beta l \left( x_b \frac{dy}{d\lambda} - \gamma \frac{dx_a}{d\lambda} \right) = 0 \quad (41)$$

Substituting Eqs. (21), (22) and (23) into Eqs. (40) and (41), one obtains

$$\frac{dx_a}{d\lambda} = \frac{4\theta f \tau_b \left( 2\tau_b f^2 \frac{\partial y}{\partial f} + \alpha y \right)}{\gamma \left( 8f^3 \tau_b^2 \frac{\partial y}{\partial f} + 8\alpha y f \tau_b - \alpha^2 \gamma \right)}. \quad (42)$$

Eq. (42) captures the change of modal split under the combined impact of space allocation and bus service adjustments. Apparently, Eq. (42) implies that expanding bus lanes does not necessarily promote bus ridership in the bimodal system. Substituting Eq. (42) into Eq. (39), one obtains the marginal effect of road space allocation on the system cost.

**Proposition 4..** Under the profit-maximizing bus service frequency and fare established in Proposition 2, expanding dedicated bus lanes reduces the system cost if and only if

$$(i) \frac{dx_a}{d\lambda} < -\frac{2n_c \theta x_b v_a^2 (v_a - v_c)}{Dlv_c} - \frac{x_a}{1-\lambda}, \text{ where } 0 < n_a \leq n_c; \quad (43)$$

$$(ii) \frac{dx_a}{d\lambda} > \frac{l\theta x_a^2 x_b}{Dn_j} - \frac{x_a}{1-\lambda}, \text{ where, } n_c < n_a \leq n_j. \quad (44)$$

**Proof..** Substituting Eqs. (23), and (27) into Eq. (38), we have

$$\begin{aligned} \frac{dSC}{d\lambda} &< 0 \\ \Leftrightarrow \begin{cases} \frac{dx_a}{d\lambda} < -\frac{2n_c \theta x_b v_a^2 (v_a - v_c)}{Dlv_c} - \frac{x_a}{1-\lambda}, & \text{if } 0 < n_a \leq n_c \\ \frac{dx_a}{d\lambda} > \frac{l\theta x_a^2 x_b}{Dn_j} - \frac{x_a}{1-\lambda}, & \text{if } n_c < n_a < n_j \end{cases}. \square \end{aligned}$$

Proposition 4 establishes the analytical conditions with which expanding bus lanes is beneficial to the whole system when the bus service parameters are responsive. If a marginal increase of the space fraction effectively reduces the private car use while keeping the traffic condition on the left branch of the MFD, it also reduces the system cost. When the car traffic is hypercongested, the system cost can be reduced only if the increase of space fraction attracts more private car users. The divergent properties on the two branches of the MFD stems from the distinct effect of the modal split on traffic speed obtained in Lemma 1. When traffic is in uncongested regime ( $0 < n_a \leq n_c$ ), larger car inflow results in lower speed, whilst in hypercongestion ( $n_c < n_a \leq n_j$ ) larger inflow corresponds to higher speed.

The observations in Proposition 3 and Proposition 4 are combined to give the following proposition.

**Proposition 5..** A Pareto-improving win-win situation (where an expansion of the bus lane results in both bus profit increase and system cost reduction) occurs if and only if either Condition (43) holds or the intersection of Conditions (26) and (44) is valid.

Proposition 5 establishes the conditions where expanding bus lanes is beneficial to both the bus operator and the system as a whole in the bimodal system with space separation.

**Table 1**  
Functional forms and parameter values.

	Functional forms	Parameter values
Car traffic speed-density relation	Eq. (7)	$v_c = 40\text{km/h}$ $k_j = 150\text{veh/km}$ $k_c = 30\text{veh/km}$
Bus speed	$v_b(\lambda) = v_{b0} - \delta(\lambda)x_b, \delta(\lambda) = \delta_0 - \delta_1\lambda$	$v_{b0} = 60\text{km/h}$ $\delta_0 = 0.003$ $\delta_1 = 0.001$
In-vehicle travel time	$tt_i = \frac{l}{v_i}, i = a, b$	$l = 12\text{km}$
Full prices	Eqs. (9) and (11)	$\tau_a = 1.5\$/\text{trip}$ $\beta = 10\$/\text{h}$ $\alpha = 1.7\beta$
Bus operating cost	$K(f) = K_0 + K_1f$	$K_0 = 1, 000\$/\text{h}$ $K_1 = 10\$/\text{run}$

## 6. The system performance under static space distribution: a numerical illustration

This section employs numerical analysis to illustrate the quantitative effects of static road space distribution. In contrast to Section 7 where the dynamic space distribution will come into play, everything shown in this section is the steady-state result – upon any snapshot, the input variables (e.g., road space allocation, travel demand) are static and the output variables (traffic status, bus service, modal split, etc.) are in equilibrium state.

Consider a downtown area with the radius of 5 km and the length of 150 km of vehicle lanes. A fraction of road space is dedicated to buses only. We assume that the travel demand is evenly generated in the region, and the congestion of car traffic is homogeneously distributed over space and exhibits an MFD. The function forms and parameter values for the numerical analysis are listed in Table 1. In particular, the marginal effect of travel demand on average bus speed  $\delta$  is represented by a linear function of the space allocation.<sup>4</sup> The function forms and parameter values used in the numerical analysis are for illustrative purpose. They should receive careful calibration and evaluation when applying the proposed model for different cities in order to draw specific conclusions and suggestions.

In contrast to the earlier analytics that are limited to marginal changes, the numerical analysis in this section is conducted over a wide range of the space allocation choices and/or demand levels.

Fig. 2 shows the effects of increasing total travel demand in the bimodal system when a constant fraction  $\lambda = 0.2$  of the total road space is allocated for bus lanes. While the demand changes, the bus fare and service frequency are adjusted accordingly by the profit-maximizing bus operator. It is found that the density of cars rises to the critical density  $k_c$  when the demand approaches  $1.62 \times 10^4\text{prs/hr}$ ; above this critical demand level, sharp upsurges occur in car accumulation and bus fare with downturn in service frequency when car traffic enters the hypercongested regime. Such drastic changes lead to the incomparable scales of variables in the two regimes of the MFD. For illustration purpose, the results are thus displayed in two separated ranges of the total demand.

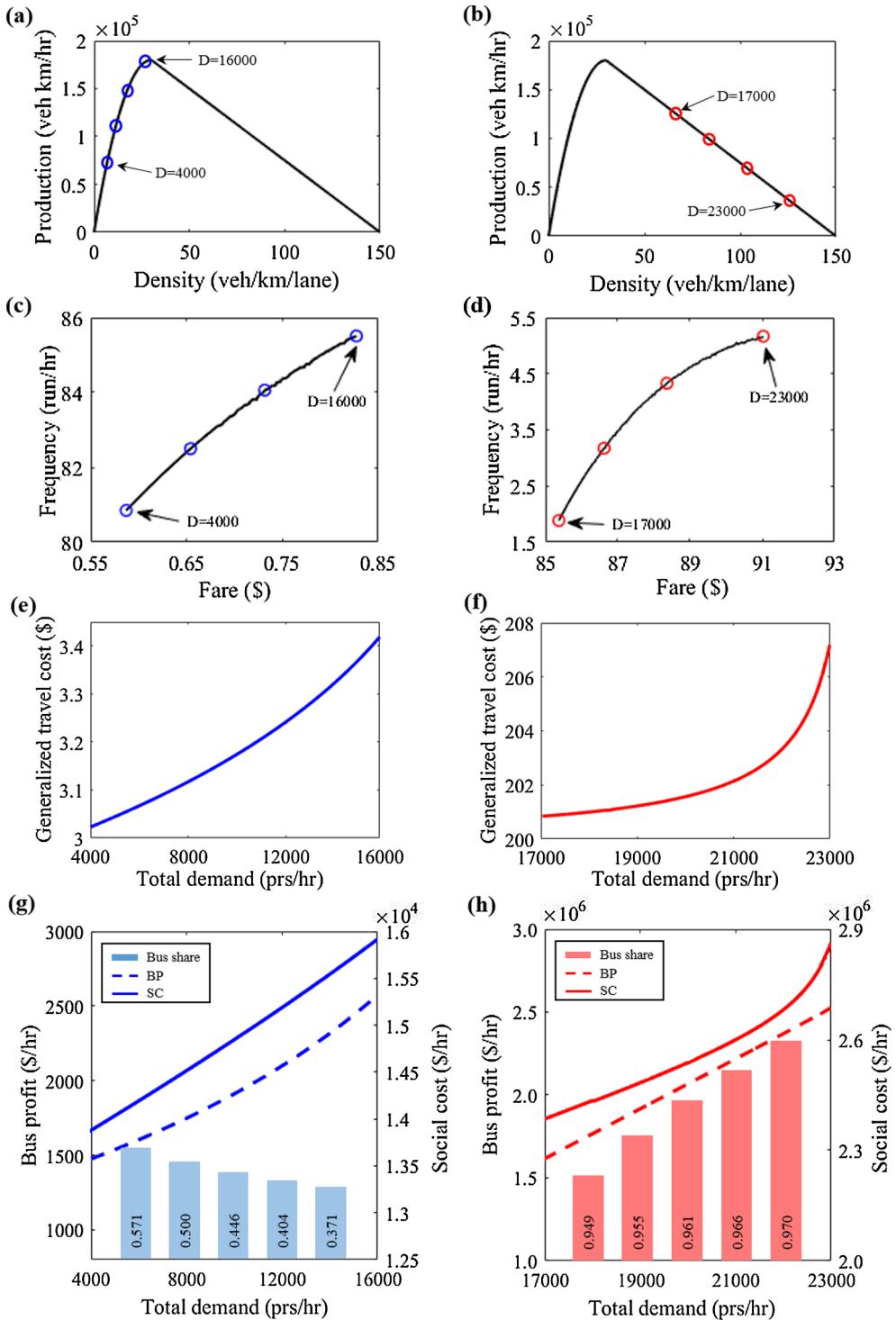
The left panel of Fig. 2 shows the results where the demand ranges from  $4 \times 10^3\text{prs/hr}$  to  $1.6 \times 10^4\text{prs/hr}$  and the right panel the range of  $1.7 \times 10^4\text{prs/hr}$  to  $2.3 \times 10^4\text{prs/hr}$ . It is observed from the figure that increasing total demand has similar overall impacts in both regimes of the MFD in terms of the change directions of car traffic density, bus fare, service frequency,<sup>5</sup> individual travel cost, operator's profit and social cost. When the total demand increases, cars accumulate (Fig. 2(a) and (b)), bus fare rises, service gets denser (Fig. 2(c) and (d)), travel cost escalates (Fig. 2(e) and (f)), and bus operator earns more causing larger deadweight loss (social cost) to the system as a whole (Fig. 2(g) and (h)).

The key difference of the two regimes lies in the change of the mode share as shown in Fig. 2(g) and (h). In the uncongested regime, increasing total demand leads to the reduction in the mode share of bus. The reduction is given by the relatively larger marginal bus cost associated with the modal split than the marginal driving cost. In the current situation, production rate of the car traffic increases with the demand as shown in Fig. 2(a) and the car system is able to accommodate more demand at a relatively lower cost. The changes of frequency and fare however result in larger cost in accommodating demand increments. It is noteworthy that the mode share is merely a relative term when the total demand changes – reduction in mode share does not imply the reduction in absolute number of users, nor the reduction in profit. As shown in Fig. 2(g), even though the bus mode share decreases, neither the number of bus users nor the bus profit decreases with the demand level.

When the car traffic is in hypercongestion, the mode share of bus surges to above 90% and keeps increasing with the total demand. Even though the bus fare has been extraordinarily high and frequency extremely low, most users choose bus mode because choosing the alternative means running at a crawling speed in hypercongestion. Bus looks almost like a monopoly which allows the

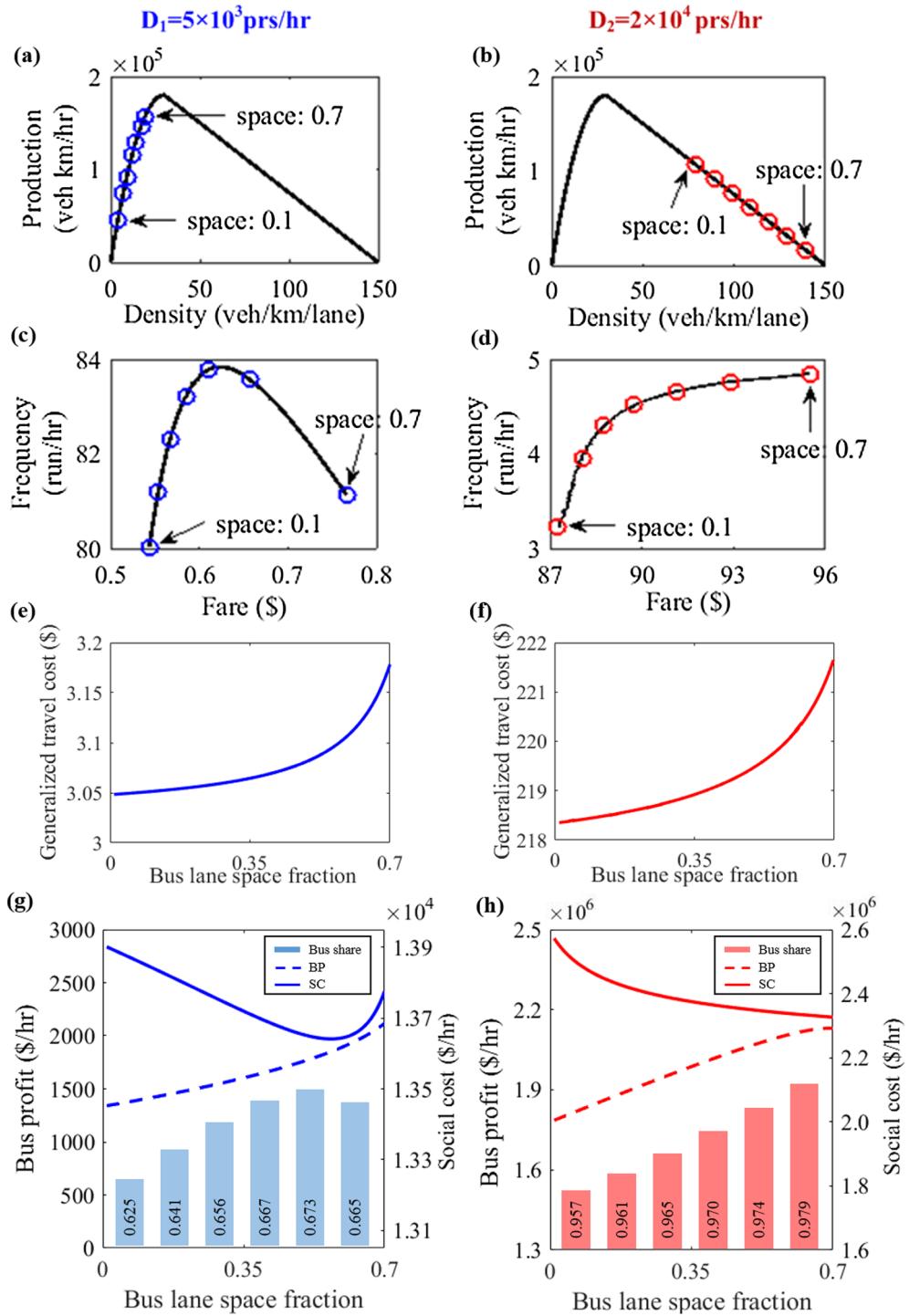
<sup>4</sup> The analysis in Sections 3–5 does not rely on the linearity; instead, it only requires the effect decreases with the space allocation, i.e.,  $\delta' = d\delta/d\lambda < 0$ .

<sup>5</sup> The bus traffic here is an aggregation of the traffic flow of all buses running within dedicated bus lanes. Such aggregated traffic contains multiple bus lines that share similar routes/segments. Thus the service frequency reflects the aggregated frequency of all buses.



**Fig. 2.** The effects of total demand on car traffic, bus service, individual travel cost, modal split, bus profit, and social cost with profit-maximizing bus service.

bus operator to charge extremely high fares and maximize its own profit. Nevertheless, this stems largely from the assumption that the total demand for cars and buses is perfectly inelastic to travel cost. In reality, users have other choices (e.g., not to travel, cycling) and equilibrium will be at a different point, probably resuming the uncongested regime. Also, our analysis ignores that bus capacity is restricted, which will most probably create another equilibrium point (by influencing the actual waiting time in Eq. (19)).



**Fig. 3.** The effects of space allocation on car traffic, bus service, individual travel cost, modal split, bus profit, and social cost with profit-maximizing bus service.

Fig. 3 shows the effects of increasing the proportion of road space allocated for bus lanes with constant total demand. In the meantime, the bus fare and frequency are governed by the profit-maximizing bus operator. The left panel of Fig. 3 presents the low demand case with  $D_1 = 5 \times 10^3$  prs/hr, and right panel the high demand case with  $D_2 = 2 \times 10^4$  prs/hr. Considering that in reality it is impossible to indefinitely expand the bus lane, we set an upper-limit for the space fraction ( $\bar{\lambda} = 70\%$ ). We expect the upper-limit would be lower than 70% in practice. The purpose of choosing such a value is to present as many numerical results as possible. Under

each demand level, the fraction of bus lane space is increased from 10% to 70%.

There are several common results under both demand levels. Firstly, the density of car traffic increases when the space for normal traffic is shrinking in both Fig. 3(a) and (b). The increase in density leads to the reduction in the traffic speed and the rise in the travel cost by car. Correspondingly we observe increasing equilibrium travel cost in both Fig. 3(e) and (f). Secondly, we observe that the bus frequency changes in the same direction with the bus mode share in Fig. 3(c) and (g), or in Fig. 3(d) and (h). This is in line with the result of Proposition 2. Thirdly, with either demand level, the bus operator obtains more profit when more space is allocated for bus lanes and the car traffic gets more congested. Fourthly, it can be inferred from Fig. 3(g) and (h) that the total social cost is negatively correlated with the bus mode share – when bus mode share rises total social cost drops, and vice versa. In the low demand case, the minimal social cost is achieved when the bus mode share is maximized and the space fraction is 0.568. The total social cost could be reduced by allocating more space to buses when the car traffic is hypercongested. Therefore, Fig. 3(g) and (h) suggest that increasing bus lane space is generally beneficial for the monopoly bus operator, and that it is socially preferable to improve the mode share of bus in the bimodal system.

Comparing the results of different demand levels, we could see that in the low demand case, the car traffic is in the uncongested regime and the bus operator provides more frequent service with lower fare. In the high demand case, the car traffic is hypercongested. Over 90% of the travelers choose the bus mode, even though the bus operator provides sparse service with high fare. Under low demand, the marginal effect of space allocation on bus mode share decays with the space fraction if car traffic remains in the uncongested regime (refer to Fig. 3(a) and (g)). But in the high demand case, the bus mode share increases more sharply when the space fraction is high (refer to Fig. 3(h)).

Similar concerns for the hypercongested regime to the analysis of Fig. 2 remain here. While the hypercongested results are included here for the sake of completeness, it is not expected that hypercongestion coexists with a stable bimodal equilibrium for a long time. As we will describe later, hypercongestion yields more interesting results with time-varying space allocation.

Normally the allocation of road space is decided by the transport authority at the higher level. While the transport authority is concerned with social welfare when shaping the space allocation, the private bus company is interested in its own profit when determining the service plan and ticket price. The above has shown that the two objectives (social cost minimization and bus profit maximization) are partially conflicting. Therefore, it is necessary for the transport authority to take bus operator's responses into account.

Fig. 4 presents the interaction between bus profit and social cost in the two dimensional space. The x-axis represents the space fraction for bus lanes and the y-axis represents the total travel demand. For each combination of space allocation and travel demand, bus operator chooses the optimal frequency and fare to maximize its profit (as in Proposition 2). Figs. 2 and 3 have shown that when total demand is high the car traffic would surge to hypercongestion. In Fig. 4 the boundary where the car traffic reaches hypercongestion is outlined by the dash-dotted line. It is evident that the critical demand is lower if less road space is reserved for car traffic (from left to right). The area above the critical demand is dyed in grey highlighting the hypercongested car traffic. In each area, the blue solid curves sketch the bus profit contours led by the optimal frequency and fare. The red dashed curves are the resulting

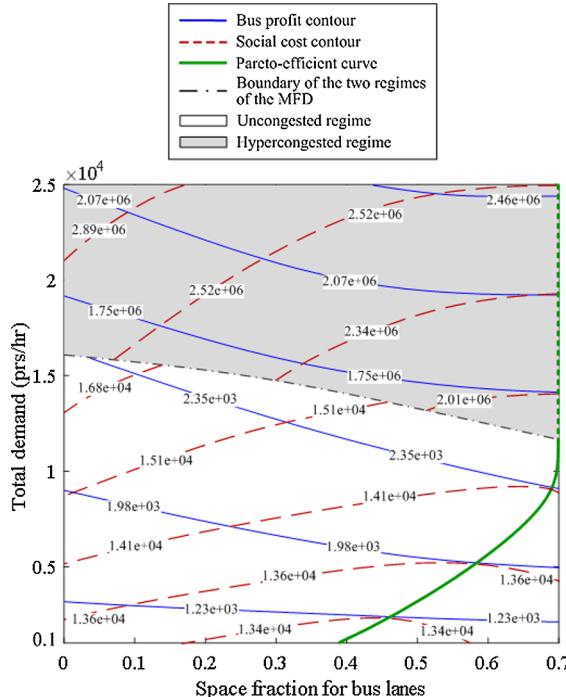


Fig. 4. Contour map of social cost and bus profit in the domain of space fraction and total travel demand.

**Table 2**

The Pareto-optimal space fraction under profit-maximizing bus service.

Total demand ( $\times 10^4$ prs/hr)	0.25	0.5	0.75	1	>1.17
Pareto-optimal space fraction	0.491	0.568	0.616	0.695	$\bar{\lambda}$

social cost contours. The green curve outlines the Pareto-efficient frontier in the uncongested regime where neither of the two objectives can be further improved without reducing the other. In the hypercongested regime the Pareto-efficient curve does not exist. For given demand level, the space allocation that simultaneously maximizes bus profit and minimizes social cost is the upper-limit. Thus, it is always preferable to allocate as more space to bus lanes as possible pushing the optimal space allocation to the upper-limit (shown in green dotted line). This reflects the advantage of giving priority to public transport during the peak period.

For given levels of total demand, the Pareto-optimal space fractions resulting the minimal social cost and maximal bus profit are summarized in [Table 2](#). It is generally suggested that larger space should be allocated for buses when total demand increases. In practice the travel demand fluctuates over time within a day. It is relatively low in off-peak periods while much higher in peak periods. This requires the space allocation to be responsive to the demand changes, which motivates the investigation of the dynamic space allocation in the next section.

## 7. The dynamic space allocation and responsive bus service

In this section, we carry out numerical analysis on the bus service design in response to the dynamic space allocation, and its impact on the traffic performance. Our motivation is to develop and evaluate different operation strategies, which are typically complex to obtain analytically. The discussion based on simulation study will provide insights into multimodal planning and management.

### 7.1. Case study set-up

Consider a downtown area with the radius of 5 km (same as in the static case). A fraction of road space is dedicated to buses only. We simulate an urban road traffic system for 4 h ( $T = 80$  time units, 3 min each), a typical morning or evening period. Demand has a symmetric trapezoidal shape with time and the length of peak period is equal to 1 h, shown in [Fig. 5\(a\)](#).

The setting of the MFD for car traffic follows: the free flow speed for cars  $v_f = 32\text{km/h}$ , critical density  $k_c = 0.03\text{veh/m}$ , jam density,  $k_j = 0.15\text{veh/m}$ . For bus running speed  $v_b(t; \lambda) = v_{b0} - \delta(\lambda)x_b(t)$ ,  $v_{b0} = 0.75v_f$ . The coefficient  $\delta$  is set as a decreasing linear function of the space allocated to buses  $\lambda$ :  $\delta(\lambda) = \delta_0 - \delta_1\lambda$ , with  $\delta_0 = 0.02$ ,  $\delta_1 = 0.015$ .

### 7.2. Dynamic space allocation

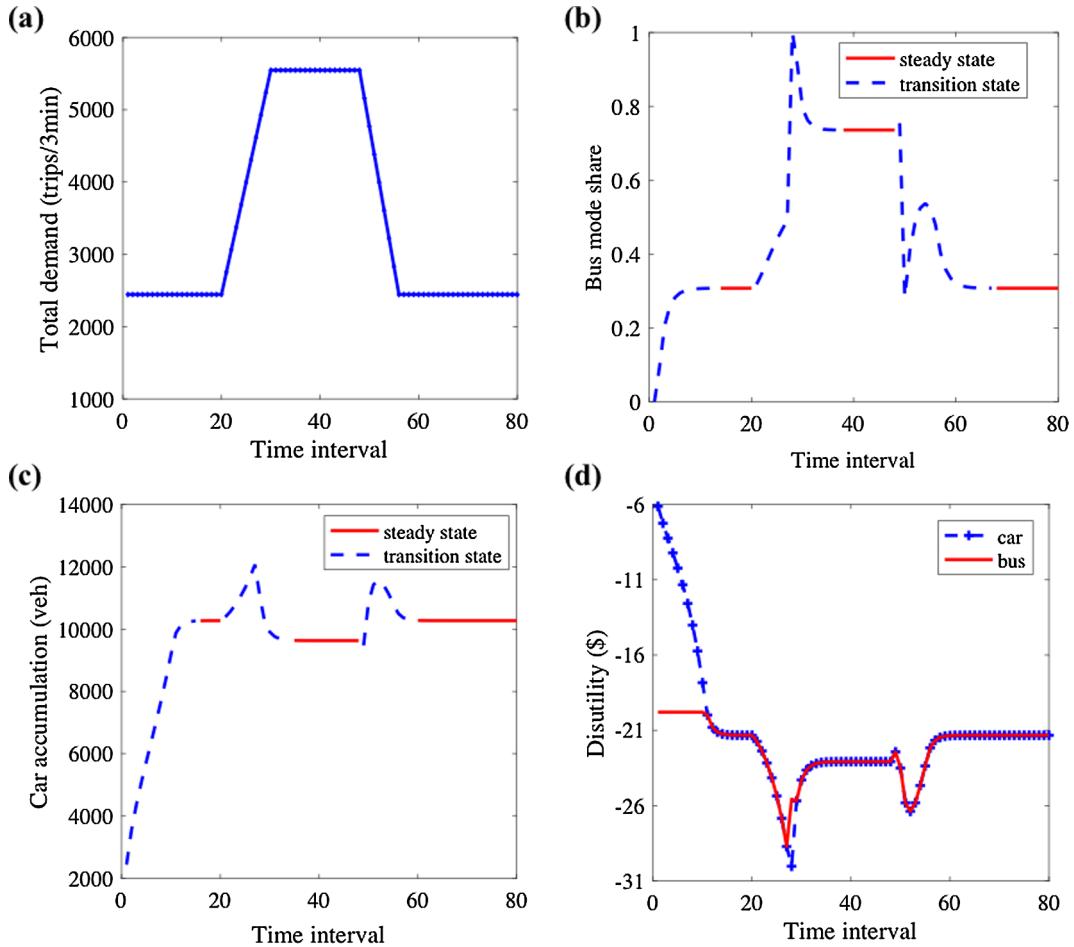
We consider a time-dependent space allocation strategy: in off-peak periods,  $0 < t \leq t_s$  and  $t_e < t < T$ ,  $\lambda = \lambda_1$ ; during the peak period,  $\lambda = \lambda_2 \geq \lambda_1$ . Thus  $t_s$  and  $t_e$  represent the starting and ending time of the peak-hour space allocation, and  $0 < t_s < t_e < T$ . As travelers' reaction to the change of space allocation would boost travel mode share during the peak period, the service frequency and fare should be adjusted during the peak period, in order to maintain service level and efficiency of the space. We thus propose the responsive frequency and fare settings in the following way: when  $\lambda = \lambda_1$ ,  $f = f_1$  and  $\tau_b = \tau_{b1}$ ; when  $\lambda = \lambda_2$ ,  $f = f_2$  and  $\tau_b = \tau_{b2}$ .

Now we examine scenarios where the demand-responsive space allocation and bus service are implemented where  $\lambda_1 = 4\%$ ,  $\lambda_2 = 14\%$ ,  $\tau_{b1} = 1\$$ ,  $\tau_{b2} = 3\$$ ,  $f_1 = 0.2\text{run/min}$ ,  $f_2 = 1\text{run/min}$ ,  $t_s = 28$ , and  $t_e = 50$ . We present in [Fig. 5](#) the evolutions of (a) total demand, and the resultant (b) mode share, (c) car accumulation and (d) travel disutilities (costs) over time. [Fig. 5](#) demonstrates that the steady state is achieved in each of the implementing time intervals:  $0 \sim t_s$ ,  $t_s \sim t_e$ ,  $t_e \sim T$ . During each steady state, the car accumulation is constant and travel disutilities are identical for the two modes. Under such steady state, the dynamic user equilibrium is reached and the resulting mode share is exactly equal to the static bimodal equilibrium determined by Eq. (12). This indicates that the steady-state analysis in [Sections 3–6](#) also has implications in the dynamic context.

[Fig. 5](#) also shows that the variation of mode split closely follows that of disutilities which reflects consistency. Car accumulation, as a result of mode choice, shares the trend in time and explains the evolution of disutility. In between the steady states, transition periods exist where the traffic status and modal split fluctuate when the space allocation and bus service changes. This is driven by how we increase or decrease the space fraction which fundamentally changes the traffic density and the associated travel costs.

### 7.3. Optimal responsive bus services

The efficiency of space allocation strategies in congestion management can be strongly influenced by the level of transit operation. If a large amount of space is dedicated to buses while the service frequency is limited or bus fare is overcharged, travelers may not choose buses. Congestion remains as mode shift is insufficient and space will be wasted. On the other hand, if many buses are deployed in service with low fare charged, transit operators may lose profit due to high operational cost. Furthermore, if more buses are present in the network and larger mode share is requested, the service level of bus is also impeded as the operating speed of buses decreases (recall from Eq. (10)), resulting longer travel time.



**Fig. 5.** (a) Total travel demand for the bi-modal system, (b) bus mode share, (c) car accumulation, and (d) travel disutilities under dynamic space allocation and bus service.

It can be seen that the bus operation optimization problem is challenging because different authorities are involved with different concerns and preferences when planning infrastructure, setting up service frequencies and bus fares. This problem is even more complex when dealing with multimodal networks, where the impact of bus operation policies and strategies on individual modes should be considered. To this end, we carry out an optimization framework to obtain optimal service frequencies and bus fares from three perspectives: total travel cost of the travelers (TTC), total profit of the bus operators (TBP), and total system cost (TSC).

The TTC includes travel time, waiting time (for buses), fixed cost of using a mode and the total money paid on bus fare. The TBP contains two elements, the total money paid on bus fare minus the cost of operating a certain frequency of service. Under this treatment, optimizing the TTC (by minimization) can have a direct conflict with optimizing the TBP (by maximization). By definition, the TSC is the sum of travel costs and bus operational costs, subtracting the total money paid on bus fare (as this amount is both cost and benefit). Optimizing TSC then aims to achieve a global minimization of travel time of the users and bus operational cost of the operators.

Optimization of the aforementioned three problems is highly non-linear. We solve this problem by a non-linear programming method, the sequential quadratic programming (SQP), which solves a sequence of optimization sub-problems (each of which optimizes a quadratic model of the objective subject to a linearization of the constraints). For detailed description of the SQP algorithm, please refer to Nocedal and Wright (2006). We apply this algorithm for multiple initial values (around 100) to avoid convergence to local minima, which might be the case for a non-smooth objective function.

Table 3 summarizes the optimal frequency, bus share, and the resultant total travel cost (TTC), total bus profit (TBP), and total system cost (TSC) under the three optimization scenarios. To achieve the optimal TTC, it is found that frequent bus service and cheap bus fare are required. This is a logical result, as operation cost is not taken into account. A policy-maker can simply decide to deploy as many buses as possible to decrease the waiting time of bus users, in order to lower the TTC. Such policy would not be in favor for the bus operators, as the resultant TBP has a negative value indicating that the operational cost exceeds the profit from bus fare collection.

Comparing to the optimal frequencies and bus fares of the TBP, the picture is completely different. An interesting observation, as

**Table 3**

Comparison of performances under different optimal cases.

Bus operator's objective	The optimal frequency (bus/10 min)		The optimal fare (\$/trip)		TTC (10e5 min)	TBP (10e4 \$)	TSC (10e5 min)
	$f_1$	$f_2$	$\tau_{b1}$	$\tau_{b2}$			
Min TTC	5	10	0.1	0.5	<b>6.65</b>	-4.03	7.13
Max TBP	0.65	1	3.6	5	14.3	<b>3.17</b>	14.9
Min TSC	0.5	7.5	0.1	0.5	6.77	2	<b>7.03</b>

The numbers in bold-italic style indicate that they are the smallest values in the relative columns.

displayed in Fig. 6(a), is that the mode share of buses is zero until the middle of the peak hour (interval 40). The high price of the bus fare, 3.6\$ per trip, makes the bus a costly mode to be utilized in the beginning of the simulation period. It may seem counter-intuitive that the bus operator sets up such price that drives away the users (customers). Let us now think a step further based on the resultant MFDs illustrated in Fig. 6(b). We will understand that this is an excellent strategy for maximizing the TBP. In Fig. 6(b), it is clearly shown by the MFD (the blue scatters) that network under a TBP operation experiences significant congestion, represented by the decreasing part of the curve. Note that there are two MFDs for each strategy (for strategy TBP it is more noticeable), each MFD corresponds to one space allocation ( $\lambda, 1-\lambda$ ) which offers different capacities for cars. As the car network becomes more and more congested and the travel cost by bus outperforms the one by car, the share of bus users is increasing, even though the bus fare increases to 5\$ per trip. This is a typical consequence of monopoly, where the bus operator enforces a profitable price given that users have no other alternative. The resultant TBP outweighs the other two cases; however, TTC and TSC are twice as much as the ones of the other two cases.

The last strategy aims to minimize the total system cost TSC. The main difference between this strategy and the strategy to optimize TTC lies in the inclusion of the operational cost. As a result, we can see in Table 3 that (i) the optimal frequencies in this case are smaller and more reasonable than the ones that optimize TTC, and (ii) the resultant TBP in this case is not only positive but also reaches a decent amount. Fig. 7 below compares the time series of disutility by mode (travel cost in negative value) and travel time by mode. In Fig. 7(a)–(c), equilibria of travel cost between the two modes can be found in all the optimal conditions of the three strategies, albeit the duration of the equilibrium is uneven. Fig. 7(d)–(e) plot the travel time (TT) time series of cars and buses. For the optimal TSC, it is promising to see that the disutility of bus improves during peak-hour as travel time decreases. This accounts for a nearly 30% mode shift (see in Fig. 6(a) the curve in green).

As a general remark, the results in this section highlight the difference in performance among bus operations of different objectives and strategies under dynamic space allocation. We demonstrate that strategies that optimize the TTC or TBP ignore important costs from the other's perspective. The strategy resulting in the lowest TSC shows strong advantage in global optimum, though bus or transit operators may need to negotiate with the traffic operators to obtain a higher TBP at an acceptable additional expense on the TTC.

#### 7.4. The importance of space allocation

Let us consider space share for buses is not optimal, for example applying a 10% constant space over time, and reproduce the results as in Table 4. System cost and social cost increase regardless of reaching which objective; while the revenue for bus operator decreases. Comparing to Table 3, it is found that the difference in costs or revenue is between 20% and 25% (e.g. from 7.03 to 8.62, from 3.17 to 2.40). When optimizing bus revenue, service frequency is adjusted to the change of space. In this case, more service is provided during off-peak in response to more space; while less frequent during peak-hour to reduce operation cost, compensating the lost revenue due to its lower speed. Similar trend can be observed when optimizing system cost. Interestingly, bus service frequency

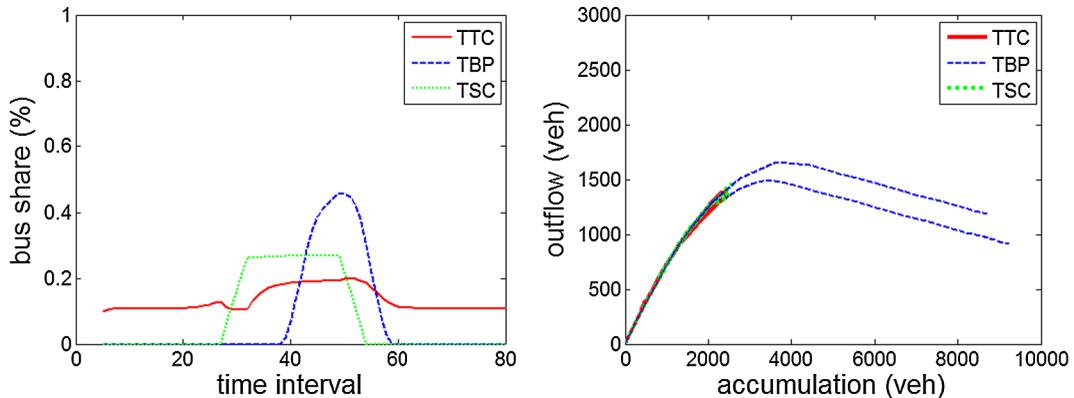
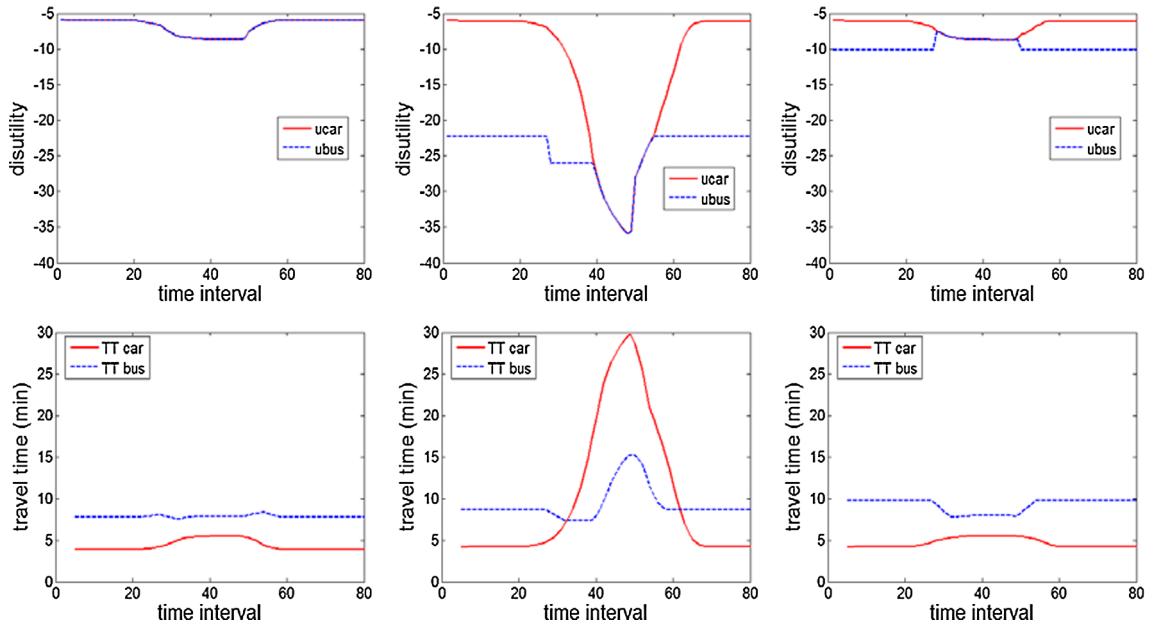


Fig. 6. The resultant (a) mode share of buses and (b) MFD of the three strategies.



**Fig. 7.** Disutility time series for using cars and buses under the three strategies: (a-top left) TTC, (b-top medium) TBP, and (c-top right) TSC; Travel time series (d-bottom left) TTC, (e-bottom medium) TBP, and (f-bottom right) TSC.

**Table 4**

Comparison of performances under different optimal cases and constant space allocation (non-optimal).

	The optimal frequency (bus/10 min)		The optimal fare (\$/trip)		TTC (10e5 min)	TBP (10e4 \$)	TSC (10e5 min)
	$f_1$	$f_2$	$\tau_{b1}$	$\tau_{b2}$			
Min TTC	5	10	0.1	0.5	<b>8.24</b>	-3.65	8.80
Max TBP	1	0.8	4.5	5	15.3	<b>2.40</b>	15.8
Min TSC	2	4.3	0.1	0.5	8.31	1.73	<b>8.62</b>

during peak-hour drops almost 40% (from 7.5 to 4.3), as a response to less lane space. If space is properly allocated, deploying more buses can actually benefit the entire system (smaller TSC) as well as the bus operator (higher TBP), which highlights the importance of space optimization.

## 8. Conclusion and discussion

This paper investigates the system-wide effect of the road space distribution in a bimodal (private and public transportation modes) system and how the bus operator responds to the space changes. The bimodal dynamic user equilibrium is formulated for separated road space in the framework where the MFD model captures the dynamics of the car traffic and the delay of a bus trip depends on both space allocation and travel demand.

The profit-maximizing bus fare and service frequency are firstly recognized for given space allocation in the steady state. In particular the counter-intuitive properties of traffic equilibrium and optimal bus service design in hypercongestion have been discovered. It is shown that when car traffic is in the uncongested regime the car usage increases if bus service level decreases. When the car traffic is in hypercongestion, higher usage corresponds to higher speed. As the consequence, neither improving bus service nor providing more space for buses could necessarily increase the bus ridership. The optimal frequency is in line with the “square root principle” implying the profit-maximizing bus service also exhibits increasing returns to scale. It is further revealed that under comparable conditions, the service frequency would be identical on the two branches of the MFD, while the bus fare would be higher in the uncongested regime than in hypercongestion. To understand the interaction between the transport authority (system benefit-maximization) and the bus operator (profit-maximization), we examine the effect of road space allocation on the bus operator’s benefit and explore how the total system benefit changes with the allocation. The marginal profit of bus operation associated with marginal bus lane capacity is proportional to the bus fare, while its change direction is consistent with the marginal bus ridership. With the responsive bus service, the condition under which expanding bus lane capacity is beneficial to the system as a whole has been established. Numerical analysis shows that the total social cost can be reduced by improving the mode share of bus service where the service is determined by a profit-maximizing operator. A Pareto-optimal space allocation that simultaneously minimizes social cost and maximizes bus profit has been identified for given total travel demand. It is generally suggested that larger space

should be allocated for buses when total demand increases. This requires the space allocation to be responsive to demand changes in the temporal horizon. The model is thus applied to the dynamic framework where the space distribution changes with the varying demand and a case study is carried out on the responsive bus service to the dynamic space distribution. We have explored the optimal bus service in terms of system cost-minimization, bus profit-maximization and user cost-minimization. The effectiveness and in congestion management of the three operation strategies are compared. Results show that different space allocation and bus service strategies can be developed utilizing the proposed system model and optimization.

This paper integrates the bimodal equilibrium model and the aggregate traffic flow model. In contrast with other models of traffic congestion, e.g., the BPR function where congestion is assumed to depend only on the current demand without any memory for the level of congestion at previous times, the MFD captures the state-dependent hypercongestion where flow decreases as density increases beyond a critical value of density, and is consistent with the physics of traffic congestion at network level. The integration with MFD models enables the model to describe the aggregate traffic status using a well-defined relationship while allows for analytical tractability to some extent. In this sense, this paper contributes to the literature by providing an analytical framework to model the multimodal interaction and spatial effect and to capture the aggregate properties of traffic flow, and sheds lights on the relevant policy and decision makings.

The analysis of this paper could be fruitfully extended in several directions. Firstly, the analytical derivation in this paper focuses on the properties of the profit-maximizing bus service. The comparison of bus services with other economic objectives (e.g., total system cost minimization, maintaining breakeven) deserve research efforts. Secondly, it has been implicitly assumed that the transit operator adjusts service immediately after the road space distribution is changed. In practice, such adjustments can take considerable time. To perform an accurate cost-benefit analysis, the accounting time horizon could be divided into two intervals: a short run in which transit service is fixed, and a long run in which it is endogenous. A distinction could also be made between the time required for fare adjustments and the (typically longer) time required to adjust frequency and other dimensions of capacity. Such an analysis would also be of interest in determining whether transit adjustments are beneficial or counterproductive. As an extension of the current model, the relationship between the space allocation and the bus service accessibility can be incorporated by adding an endogenous bus access cost term in Eq. (11). Theoretically, this term should negatively correlate with the space allocation parameter  $\lambda$ . In order to obtain the specific function that can well demonstrate the underlying relationship, quantifications need to be carried out based on empirical data and then can feed into the current modeling framework. The inclusion of such endogeneity will increase the complexity of the analytics, but we expect the whole analytical process is achievable following the framework of the current study and adopting similar techniques. We will actively explore this extended model in future studies.

Another interesting direction is to investigate how the road space management coupled with other strategies such as perimeter control (e.g. Ampountolas et al., 2017; Ramezani et al., 2015; Kouvelas et al., 2017), car usage tolling (e.g. Zheng et al., 2012; Daganzo and Lehe, 2015; Chen et al., 2016), and transit subsidy (e.g. Basso and Jara-Díaz, 2012; Zheng et al., 2016) can further improve the state of the multimodal system. Our results reflect clearly the conflict between the objectives between the system and bus operators (i.e. Min TSC versus Max TBP). The system operator can influence traffic performance through its controlling over the allocation of road space between general and bus lanes. The space share can be optimized, after evaluating the total social costs under a certain level of bus service. In turn, the bus operator can adjust the service frequency and fare in response to the given bus lane capacity. The nature of this relationship between the two operators, makes their interacting process a leader-follower Stackelberg Game. We are intrigued to investigate if there is an equilibrium exists between their objectives TBP and TSC through rounds of competition. Our on-going effort is towards this direction.

## Acknowledgements

The first author was financially supported by the Overseas Research Award provided by the Hong Kong University of Science and Technology while visiting École Polytechnique Fédérale de Lausanne (EPFL). This research was also supported by a grant from Hong Kong's Research Grants Council under project No. HKUST16205715, and from Chinese National Basic Science Foundation under project No. KZ73026901. Research at EPFL was partially supported by ERC Starting Grant “METAERW: Modeling and controlling traffic congestion and propagation in large-scale urban multimodal networks”. The authors are very grateful for the comments from the anonymous reviewers which significantly improve the quality of the paper.

## Appendix A. The derivations for Eqs. (17) and (18)

Eq. (16) is equivalent to

$$n_a v_a - l x_a = 0, \quad (45)$$

where  $t$  in brackets is omitted in line with the steady-state context. According to Eq. (7), where  $0 < k_a \leq k_c$  or  $0 < n_a \leq n_c$ ,

$$v_a = -\frac{v_c}{k_c} k_a + 2v_c = -\frac{v_c}{n_c} n_a + 2v_c. \quad (46)$$

Substituting Eq. (46) into Eq. (45), we have  $v_c n_a^2 - 2v_c n_c n_a + n_c l x_a = 0$ , where the unique solution in the range  $0 < n_a \leq n_c$  is

$$n_a = n_c - \frac{\sqrt{n_c v_c (n_c v_c - l x_a)}}{v_c}. \quad (47)$$

Substituting Eq. (47) into Eq. (46), we have

$$v_a = \frac{\sqrt{n_c v_c (n_c v_c - l x_a)}}{n_c} + v_c$$

Similarly, where  $k_c < k_a \leq k_j$  or  $n_c < n_a \leq n_j$ ,

$$v_a = \frac{n_c v_c}{n_j - n_c} \left( \frac{n_j}{n_a} - 1 \right). \quad (48)$$

Substituting Eq. (48) into Eq. (45), we have  $\frac{n_c v_c}{n_j - n_c} (n_j - n_a) - l x_a = 0$ , and thus

$$n_a = n_j - \frac{n_j - n_c}{n_c v_c} l x_a, \quad v_a = \frac{l x_a}{n_j - \frac{n_j - n_c}{n_c v_c} l x_a}$$

## Appendix B. The uniqueness of the interior bimodal equilibrium in the steady state

The equilibrium car usage  $x_a$  is determined by  $g(x_a) = c_b - c_a = 0$ , where  $g(x_a)$  is given in Eq. (20). Thus, the underlying assumptions for interior equilibrium are (similar assumptions are made in Zhang et al., 2014):

(i) A too congested highway if all choose to drive:

$$\begin{aligned} g(x_a = D) &< 0 \Leftrightarrow c_b < c_a \\ &\Leftrightarrow \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{\beta l}{v_{b0}} - \frac{\beta n_j}{D} + \frac{n_j - n_c}{n_c v_c} \beta l D < 0. \end{aligned} \quad (49)$$

(ii) A too congested transit if no one drives:

$$\begin{aligned} g(x_a = 0) &> 0 \Leftrightarrow c_b > c_a \\ &\Leftrightarrow \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{\beta l}{v_{b0} - \delta D} - \frac{\beta l}{2v_c} > 0. \end{aligned} \quad (50)$$

When the car traffic is on the left branch of the MFD, i.e.,  $0 < n_a \leq n_c$ ,

$$g'(x_a) = -\beta l \left[ \frac{\delta}{v_b^2} + \frac{v_c (n_c^2 v_c^2 - n_c v_c l x_a)^{-\frac{1}{2}}}{2v_a^2} \right] < 0$$

Since the interior equilibria ensure  $g(x_a = 0) > 0$ , it is evident that  $g(x_a) = 0$  has and only has one positive solution.

When the car traffic is hypercongested, i.e.,  $n_c < n_a \leq n_j$ ,

$$g(x_a) = \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{\beta l}{v_{b0} - \delta(D - x_a)} - \beta \frac{n_j - \frac{n_j - n_c}{n_c v_c} l x_a}{x_a}. \quad (51)$$

Then  $g(x_a) = 0$  is equivalent to  $h(x_a) = 0$ , where  $h(x_a)$  is the common denominator of the RHS of Eq. (51):

$$h(x_a) = \left( \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{n_j - n_c}{n_c v_c} \beta l \right) \delta x_a^2 + \left[ \left( \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{n_j - n_c}{n_c v_c} \beta l \right) (v_{b0} - \delta D) + \beta (l - n_j \delta) \right] x_a - \beta n_j (v_{b0} - \delta D),$$

which is a quadratic equation of variable  $x_a$ . When  $x_a = 0$ ,  $h(x_a = 0) = -\beta n_j (v_{b0} - \delta D) < 0$ . Under the interior equilibrium assumptions Eqs. (49) and (50), the quadratic coefficient is positive as

$$\begin{aligned} &\tau_b - \tau_a + \frac{\alpha}{2f} + \beta l \frac{n_j - n_c}{n_c v_c} = \beta \frac{n_j}{x_a} - \frac{\beta l}{v_{b0} - \delta(D - x_a)} \\ &> \frac{\beta n_j}{D} - \frac{\beta l}{v_{b0} - \delta D} \\ &> \tau_b - \tau_a + \frac{\alpha}{2f} + \frac{\beta l}{v_{b0}} + \frac{n_j - n_c}{n_c v_c} \beta l D - \frac{\beta l}{v_{b0} - \delta D} \\ &> \frac{\beta l}{2v_c} + \frac{\beta l}{v_{b0} - \delta D} + \frac{\beta l}{v_{b0}} + \frac{n_j - n_c}{n_c v_c} \beta l D - \frac{\beta l}{v_{b0} - \delta D} \\ &= \beta l \left( \frac{1}{2v_c} + \frac{1}{v_{b0}} + \frac{n_j - n_c}{n_c v_c} D \right) \\ &> 0 \end{aligned}$$

According to the laws of quadratic equations, there exists one and only one positive solution to  $h(x_a) = 0$ , and so is  $g(x_a) = 0$ .

In summary, the equilibrium demand for car  $x_a$  can be uniquely pinned down in each regime of the MFD. Given the inelastic total demand, the equilibrium mode share can thus be determined.

### Appendix C. The derivations for Eqs. (21)–(23)

Taking partial derivatives on the two sides of Eq. (19) with respect to  $f$  while treating  $x_a$  as an implicit function of  $f$ ,  $\tau_b$ , and  $\lambda$ , we have

$$\begin{aligned} -\frac{\beta l}{v_a^2} \frac{\partial v_a}{\partial x_a} \frac{\partial x_a}{\partial f} &= -\frac{\beta l}{v_b^2} \frac{\partial v_b}{\partial x_a} \frac{\partial x_a}{\partial f} - \frac{\alpha}{2f^2} \\ \Leftrightarrow \frac{\partial x_a}{\partial f} &= -\frac{\alpha}{2\beta l f^2 \left( \frac{1}{v_b^2} \frac{\partial v_b}{\partial x_a} - \frac{1}{v_a^2} \frac{\partial v_a}{\partial x_a} \right)} \end{aligned}$$

Similarly, taking partial derivatives on the two sides of Eq. (19) with respect to  $\tau_b$ , and  $\lambda$ , and with some manipulations, we have

$$\begin{aligned} \frac{\partial x_a}{\partial \tau_b} &= \frac{1}{\beta l \left( \frac{1}{v_b^2} \frac{\partial v_b}{\partial x_a} - \frac{1}{v_a^2} \frac{\partial v_a}{\partial x_a} \right)} \\ \frac{\partial x_a}{\partial \lambda} &= \frac{\frac{1}{v_a^2} \frac{\partial v_a}{\partial \lambda} - \frac{1}{v_b^2} \frac{\partial v_b}{\partial \lambda}}{\frac{1}{v_b^2} \frac{\partial v_b}{\partial x_a} - \frac{1}{v_a^2} \frac{\partial v_a}{\partial x_a}} \end{aligned}$$

By denoting  $\theta = \frac{1}{v_a^2} \frac{\partial v_a}{\partial \lambda} - \frac{1}{v_b^2} \frac{\partial v_b}{\partial \lambda}$  and  $\gamma = \frac{1}{v_b^2} \frac{\partial v_b}{\partial x_a} - \frac{1}{v_a^2} \frac{\partial v_a}{\partial x_a}$ , we obtain Eqs. (21)–(23).

### Appendix D. The Hessian matrix of the profit-maximization problem for the bus operator

Following Eqs. (21) and (22), the second-order partial derivatives of the equilibrium car usage  $x_a$  with respect to bus frequency  $f$  and fare  $\tau_b$  are:

$$\frac{\partial^2 x_a}{\partial f^2} = \frac{\alpha}{2\beta l \gamma^2 f^2} \frac{\partial \gamma}{\partial f} + \frac{\alpha}{\beta l \gamma^3}, \quad (52)$$

$$\frac{\partial^2 x_a}{\partial \tau_b^2} = -\frac{1}{\beta l \gamma^2} \frac{\partial \gamma}{\partial \tau_b} = \frac{2f^2}{\alpha \beta l \gamma^2} \frac{\partial \gamma}{\partial f}, \quad (53)$$

$$\frac{\partial^2 x_a}{\partial f \partial \tau_b} = -\frac{1}{\beta l \gamma^2} \frac{\partial \gamma}{\partial f}, \quad (54)$$

where

$$\frac{\partial \gamma}{\partial \tau_b} = -\frac{2f^2}{\alpha} \frac{\partial \gamma}{\partial f}. \quad (55)$$

Then the second-order partial derivatives of bus profit  $BP$  with respect to bus frequency  $f$  and fare  $\tau_b$  are:

$$\begin{aligned} \frac{\partial^2 BP}{\partial f^2} &= -\tau_b \frac{\partial^2 x_a}{\partial f^2} = -\frac{\alpha x_b}{f^2} \left( \frac{1}{2\gamma} \frac{\partial \gamma}{\partial f} + \frac{1}{f} \right) \\ \frac{\partial^2 BP}{\partial \tau_b^2} &= -2 \frac{\partial x_a}{\partial \tau_b} - \tau_b \frac{\partial^2 x_a}{\partial \tau_b^2} = -\frac{2}{\gamma} \left( \frac{1}{\beta l} + \frac{x_b f^2}{\alpha} \frac{\partial \gamma}{\partial f} \right) \\ \frac{\partial^2 BP}{\partial f \partial \tau_b} &= -\frac{\partial x_a}{\partial f} - \tau_b \frac{\partial^2 x_a}{\partial f \partial \tau_b} = \frac{1}{\gamma} \left( \frac{\alpha}{2\beta l f^2} + x_b \frac{\partial \gamma}{\partial f} \right) \end{aligned}$$

Thus, the Hessian matrix  $H$  of the concerned profit-maximization problem is negative semi-definite if and only if

$$H = \begin{pmatrix} \frac{\partial^2 BP}{\partial f^2} & \frac{\partial^2 BP}{\partial f \partial \tau_b} \\ \frac{\partial^2 BP}{\partial f \partial \tau_b} & \frac{\partial^2 BP}{\partial \tau_b^2} \end{pmatrix} \leq 0 \quad (56)$$

$$\Leftrightarrow \begin{cases} \frac{\partial^2 BP}{\partial f^2} = -\frac{\alpha x_b}{f^2} \left( \frac{1}{2\gamma} \frac{\partial \gamma}{\partial f} + \frac{1}{f} \right) \leq 0 \\ \frac{\partial^2 BP}{\partial \tau_b^2} = -\frac{2}{\gamma} \left( \frac{x_b f^2}{\alpha} \frac{\partial \gamma}{\partial f} + \frac{1}{\beta l} \right) \leq 0 \\ \frac{\partial^2 BP}{\partial f^2} \frac{\partial^2 BP}{\partial \tau_b^2} - \left( \frac{\partial^2 BP}{\partial f \partial \tau_b} \right)^2 = \frac{2x_b^2}{\gamma} \frac{\partial \gamma}{\partial f} + \frac{2\alpha x_b}{\beta l \gamma^3} - \frac{\alpha^2}{4\beta^2 l^2 \gamma^2 f^4} \geq 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \frac{\partial \gamma}{\partial f} \geq \max \left\{ -2\gamma \left( \frac{f^2}{\alpha x_b} + \frac{1}{f} \right), -\frac{\alpha}{x_b f^2} \left( \frac{\gamma}{2} + \frac{1}{\beta l} \right), -\frac{\alpha}{\beta^2 l^2 f^3 x_b} \left( \beta l - \frac{\alpha}{8\gamma f x_b} \right) \right\}, & \text{if } \gamma > 0 \\ \frac{\partial \gamma}{\partial f} \leq \min \left\{ -2\gamma \left( \frac{f^2}{\alpha x_b} + \frac{1}{f} \right), -\frac{\alpha}{x_b f^2} \left( \frac{\gamma}{2} + \frac{1}{\beta l} \right), -\frac{\alpha}{\beta^2 l^2 f^3 x_b} \left( \beta l - \frac{\alpha}{8\gamma f x_b} \right) \right\}, & \text{if } \gamma < 0 \end{cases} \quad (57)$$

## References

- Ahn, K., 2009. Road pricing and bus service policies. *J. Transp. Econ. Policy* 43 (1), 25–53.
- Alsalhi, R., Dixit, V., Gayah, V., 2018. On the existence of network Macroscopic Safety Diagrams: theory, simulation and empirical evidence. *PLoS ONE* 13 (8), e0200541.
- Ampountolas, K., Zheng, N., Geroliminis, N., 2017. Macroscopic modelling and robust control of bi-modal multi-region urban road networks. *Transp. Res. Part B* 104, 616–637.
- Arnott, R., 2013. A bathtub model of downtown traffic congestion. *J. Urban Econ.* 76 (4), 110–121.
- Arnott, R., Yan, A., 2000. The two-mode problem: second-best pricing and capacity. *Rev. Urban Dev. Stud.* 12, 170–199.
- Basso, L.J., Guevara, C.A., Gschwender, A., Fuster, M., 2011. Congestion pricing, transit subsidies and dedicated bus lanes: efficient and practical solutions to congestion. *Transp. Policy* 18 (5), 676–684.
- Basso, L.J., Jara-Díaz, S.R., 2012. Integrating congestion pricing, transit subsidies and mode choice. *Transp. Res. Part A* 46 (6), 890–900.
- BPR, 1964. Traffic Assignment Manual: Bureau of Public Roads. U.S. Department of Commerce, Washington, D.C.
- Chen, X.M., Xiong, C., He, X., Zhu, Z., Zhang, L., 2016. Time-of-day vehicle mileage fees for congestion mitigation and revenue generation: a simulation-based optimization method and its real-world application. *Transp. Res. Part C* 63, 71–95.
- Chiabaut, N., Xie, X., Leclercq, L., 2014. Performance analysis for different designs of a multimodal urban arterial. *Transp. B: Transp. Dynam.* 2 (3), 229–245.
- Daganzo, C.F., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. *Transp. Res. Part B* 41 (1), 49–62.
- Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2012. The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. *Euro J. Transp. Logist.* 1 (1–2), 47–65.
- Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. *Transp. Res. Part B* 42 (9), 771–781.
- Daganzo, C.F., Lehe, L., 2015. Distance-dependent congestion pricing for downtown zones. *Transp. Res. Part B* 75, 89–99.
- Eichler, M., Daganzo, C.F., 2006. Bus lanes with intermittent priority: strategy formulae and an evaluation. *Transp. Res. Part B* 40 (9), 731–744.
- Geroliminis, N., Daganzo, C., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. *Transp. Res. Part B* 42 (9), 759–770.
- Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. *Transp. Res. Part C* 42, 168–181.
- Gonzales, E.J., Daganzo, C.F., 2012. Morning commute with competing modes and distributed demand: user equilibrium, system optimum, and pricing. *Transp. Res. Part B* 46 (10), 1519–1534.
- Guler, S.I., Gayah, V.V., Menendez, M., 2016. Bus priority at signalized intersections with single-lane approaches: a novel pre-signal strategy. *Transp. Res. Part C* 63, 51–70.
- Guler, S.I., Menendez, M., 2015. Pre-signals for bus priority: basic guidelines for implementation. *Public Transp.* 7 (3), 339–354.
- Haddad, J., Ramezani, M., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. *Transp. Res. Part B* 54, 17–36.
- He, H., Guler, S.I., Menendez, M., 2016. Adaptive control algorithm to provide bus priority with a pre-signal. *Transp. Res. Part C* 64, 28–44.
- He, H., Menendez, M., Guler, I., 2017. Analytical evaluation of flexible sharing strategies on multi-modal arterials. *Transp. Res. Proc.* 23, 980–999.
- Jansson, J.O., 1984. Transport System Optimization and Pricing. Wiley.
- Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. *Transp. Res. Part B* 96, 26–45.
- Lamott, G., Geroliminis, N., 2017. The morning commute in urban areas with heterogeneous trip lengths. *Transp. Res. Part B*. <https://doi.org/10.1016/j.trb.2017.08.023>. (in press).
- Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. *Transp. Res. Part B* 57, 468–484.
- Leclercq, L., Séneçat, A., Mariotte, G., 2017. Dynamic macroscopic simulation of on-street parking search: a trip-based approach. *Transp. Res. Part B* 101, 268–282.
- Liu, W., Geroliminis, N., 2016. Modeling the morning commute for urban networks with cruising-for-parking: an MFD approach. *Transp. Res. Part B* 93, 470–494.
- Li, Z.C., Lam, W.H., Wong, S.C., 2012. Modeling intermodal equilibrium for bimodal transportation system design problems in a linear monocentric city. *Transp. Res. Part B* 46 (1), 30–49.
- Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2017. Empirics of multi-modal traffic networks – using the 3D macroscopic fundamental diagram. *Transp. Res. Part C* 82, 88–101.
- Mariotte, G., Leclercq, L., Laval, J.A., 2017. Macroscopic urban dynamics: analytical and numerical comparisons of existing models. *Transp. Res. Part B* 101, 245–267.
- Mohring, H., 1972. Optimization and scale economies in urban bus transportation. *Am. Econ. Rev.* 62 (4), 591–604.
- Nocedal, J., Wright, J., 2006. Numerical Optimization, second ed. Springer Series in Operations Research, Springer Verlag.
- de Palma, A., Lindsey, R., Monchambert, G., 2017. The economics of crowding in rail transit. *J. Urban Econ.* 101, 106–122.
- Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. *Transp. Res. Part B* 74, 1–19.
- Saberi, M., Mahmassani, H., Hou, T., Zockaei, A., 2014a. Estimating network fundamental diagrams using three-dimensional vehicle trajectories: extending Edie's definitions of traffic flow variables to networks. *Transp. Res. Rec.: J. Transp. Res. Board* 2422, 12–20.
- Saberi, M., Mahmassani, H.S., Zockaei, A., 2014b. Network capacity, traffic instability, and adaptive driving: findings from simulated urban network experiments. *Euro J. Transp. Logist.* 3 (3–4), 289–308.
- Saeedmanesh, M., Geroliminis, N., 2016. Clustering of heterogeneous networks with directional flows based on "Snake" similarities. *Transp. Res. Part B* 91, 250–269.
- Saeedmanesh, M., Geroliminis, N., 2017. Dynamic clustering and propagation of congestion in heterogeneously congested urban traffic networks. *Transp. Res. Part B* 105, 193–211.
- Tirachini, A., Hensher, D.A., 2011. Bus congestion, optimal infrastructure investment and the choice of a fare collection system in dedicated bus corridors. *Transp. Res. Part B* 45 (5), 828–844.
- Tirachini, A., Hensher, D.A., Rose, J.M., 2014. Multimodal pricing and optimal design of urban public transport: the interplay between traffic congestion and bus crowding. *Transp. Res. Part B* 61, 33–54.
- Tsekeris, T., Geroliminis, N., 2013. City size, network structure and traffic congestion. *J. Urban Econ.* 76, 1–14.
- Varian, H.R., 1992. Microeconomic Analysis. Norton & Company.
- Verbas, İ.Ö., Frei, C., Mahmassani, H.S., Chan, R., 2015. Stretching resources: sensitivity of optimal bus frequency allocation to stop-level demand elasticities. *Public Transp.* 7 (1), 1–20.
- Verbas, İ.Ö., Mahmassani, H.S., Hyland, M.F., 2016. Gap-based transit assignment algorithm with vehicle capacity constraints: simulation-based implementation and large-scale application. *Transp. Res. Part B* 93, 1–16.
- Viegas, J., Lu, B., 2001. Widening the scope for bus priority with intermittent bus lanes. *Transp. Plan. Technol.* 24 (2), 87–110.

- Viegas, J., Lu, B., 2004. The intermittent bus lane signals setting within an area. *Transp. Res. Part C* 12 (6), 453–469.
- Yang, K., Zheng, N., Menendez, M., 2017. Multi-scale perimeter control approach in a connected-vehicle environment. *Transport. Res. Part C* 94, 32–49.
- Yildirimoglu, M., Geroliminis, N., 2013. Experienced travel time prediction for congested freeways. *Transp. Res. Part B* 53, 45–63.
- Zhang, F., Yang, H., Liu, W., 2014. The Downs-Thomson Paradox with responsive transit service. *Transp. Res. Part A* 70, 244–263.
- Zhang, F., Lindsey, R., Yang, H., 2016. The Downs-Thomson paradox with imperfect mode substitutes and alternative transit administration regimes. *Transp. Res. Part B* 68, 104–127.
- Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. *Transp. Res. Part B* 57, 326–341.
- Zheng, N., Rerat, G., Geroliminis, N., 2016. Time-dependent area-based pricing for multimodal systems with heterogeneous users in an agent-based environment. *Transp. Res. Part C* 62, 133–148.
- Zheng, N., Waraich, R., Axhausen, K., Geroliminis, N., 2012. A dynamic cordon pricing scheme combining the Macroscopic Fundamental Diagram and an agent-based traffic model. *Transp. Res. Part A* 46 (8), 1291–1303.
- Zhong, R., Chen, C., Huang, Y., Sumalee, A., Lam, W., Xu, D., 2017. Robust perimeter control for two urban regions with macroscopic fundamental diagrams: a control-Lyapunov function approach. *Transp. Res. Part B* 23, 922–941.
- Zockaei, A., Saberi, M., Saeidi, R., 2018. A resource allocation problem to estimate network fundamental diagram in heterogeneous networks: Optimal locating of fixed measurement points and sampling of probe trajectories. *Transport. Res. Part C Emerg. Technol.* 86, 245–262.