On the optimization of the bus network design: An analytical approach based on the three-dimensional macroscopic fundamental diagram

Igor Dakic, Ludovic Leclercq, Monica Menendez

Transportation Research Part B 149 (2021) 393–417

Contents lists available at ScienceDirect

Transportation Research Part B

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

On the optimization of the bus network design: An analytical approach based on the three-dimensional macroscopic fundamental diagram

Igor Dakic[a,b,c,*], Ludovic Leclercq[c], Monica Menendez[d]

[a] Traffic Engineering Group, Institute for Transport Planning and Systems, ETH Zurich, Switzerland
[b] School of Civil and Environmental Engineering, Georgia Institute of Technology, USA
[c] Univ. Gustave Eiffel, Univ. Lyon, ENTPE, LICIT, Lyon, France
[d] Division of Engineering, New York University Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 25 December 2019
Revised 7 March 2021
Accepted 26 April 2021

Keywords:
Bus network design
Trip length distribution
Bi-modal MFD
Space allocation
Bus operations

ABSTRACT

Multiple factors can influence the public transport level of service. All take root in the network structure and the operating regime, i.e. how the bus lines are arranged atop the street network and how the service frequency is adjusted to meet urban mobility patterns. This is known as the bus network design problem and has been the subject of several studies. The problem is so challenging that most studies until now resort to strong assumptions such as a static description of the peak hour demand, homogeneous user behavior, and equal trip lengths. Potential effects of different types of user behavior and trip lengths patterns on the user and/or operator cost have not been investigated whatsoever. Moreover, none of the existing studies have considered the effects of bus network structure on private car users, the level of interactions between the modes, and the passenger mode choice that depends on traffic conditions.

This paper aims to close this gap and provide a general framework considering multiple trip length patterns, two types of user behavior, and the effects that the bus network structure might have on the traffic performance and passenger mode choice. For modeling different trip length patterns, we use the trip length distribution as an intermediate level of abstraction. To capture complex modal interactions and quantify the operating speeds, we apply the recently proposed three-dimensional macroscopic fundamental diagram. We use the operating speed for each mode to determine the mode choice at the trip length level. This way, we are able to solve the optimal bus network design problem under the free-flow/saturated traffic conditions in an analytical way, while considering more realistic settings including a dynamic description of the peak hour demand, mixed traffic, and different mode choice decisions depending on trip lengths and walking preferences.

Numerical analysis reveals that all the tested factors, including demand intensity, user behavior, and trip length patterns, have significant effects on the operator and user cost function. Results show that the probability of choosing any given mode follows certain distribution that varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Furthermore, the results indicate that users can benefit if they are willing to adjust the number of transfers to minimize the walking dis-

* Corresponding author.
E-mail addresses: i.dakic@ivt.baug.ethz.ch (I. Dakic), ludovic.leclercq@univ-eiffel.fr (L. Leclercq), monica.menendez@nyu.edu (M. Menendez).

https://doi.org/10.1016/j.trb.2021.04.012

tance at the origin and the destination. Moreover, we show that the optimal bus network design determined for the uniform trip pattern underestimates the number of required buses, which leads to passenger congestion at stops during the peak period. This, however, does not happen when we take into account the actual trip length distribution for the bus network design. A comparison with a simplified approach that considers the bus system only, reveals the value of accounting for the complex bi-modal interactions, especially for higher demand levels. Finally, we show that by allowing the design parameters to vary across cardinal directions we provide more flexibility for the bus system to serve the passenger demand while reducing the operator cost compared to the existing approaches.

1. Introduction and background

Multi-modality plays a critical role in mobility management. If not managed well, traffic congestion will be increasingly pervasive in urban areas. Constructing new infrastructure is an expensive solution, as the cost needed to keep pace with an increase in demand, including the induced demand, is very high (Small, 2004). Therefore, special attention should be given to the design, management, and operation of the existing road facilities for different modes. This is, however, challenging due to the complexity of the system dynamics and the lack of tools for understanding and optimizing multi-modal traffic performance under different network structures and traffic flow operations.

Motivated by the need to address the everlasting increase in travel demand, support sustainability, and preserve existing transport land use, transportation solutions are often sought in the domain of public transport systems (Roca-Riu et al., 2020). Public transport is often seen as a key component of a sustainable urban development. It is regarded as a public service that should provide mobility access to all citizens in an urban area.

One of the commonly explored problems in the public transport sector, more specifically in the domain of bus operations, is the bus network design. In this problem, the arrangement of bus lines atop a street network needs to be determined in a way that it provides a good level of accessibility and service between every pair of points in the city throughout the day. The output of this problem is used as the input to the subsequent phase of the transit network planning process (Ibarra-Rojas et al., 2015), i.e. the frequency setting problem (Dakic et al., 2021; Zhang et al., 2018; Zhang and Liu, 2019).

Several studies have looked at the optimal bus network configuration, investigating different city network structures: grid systems (Holroyd, 1967), radial systems (Byrne, 1975), systems of corridors (Wirasinghe et al., 1977), and hub-and-spoke systems (Newell, 1979). All these studies have a common objective - to determine the topological (e.g. line and stop spacings) and operational (e.g. service frequency) characteristics of the bus system that minimize the user and the operator cost. To achieve a service and accessibility level competitive with that of the automobile at a reasonable cost, Daganzo (2010) proposed a hybrid concept, obtained by combining hub-and-spoke and grid systems. This concept has further been generalized and extended to account for more realistic network configurations (Estrada et al., 2011; Badia et al., 2014; Chen et al., 2015). The model developed by Estrada et al. (2011) also served as an inspiration to construct a real transfer-based bus network, the Nova Xarxa in Barcelona, Spain.

A recent study by Badia et al. (2017) empirically proved that such a well-designed transfer-based network could attract new users, who would not be opposed to transferring. However, most scientific literature on the bus network design problem (see e.g. Daganzo, 2010; Estrada et al., 2011; Nourbakhsh and Ouyang, 2012; Badia et al., 2014; Chen et al., 2015; Amirgholy et al., 2017) assume, up to now, that bus users choose the closest origin and destination stops such that they minimize the number of transfers, adjusting their walking distance to meet this criteria. This, in turn, may result in longer access (including the egress) time, thus longer total user time traveled, depending on the selected design parameters (e.g. stop and line spacings). Potential effects of other types of user behavior (e.g. assuming that users are willing to adjust the number of transfers in order to minimize the walking distance) on the user and/or operator cost function have not been investigated.

Furthermore, these studies have also assumed trip origins and destinations to be uniformly and independently distributed across the network, reducing thereby the complexity in the mathematical modeling. Consequently, the optimal bus network design is determined for one particular trip length pattern, imposed by the homogeneously placed origins and destinations. To the best of the authors' knowledge, the only attempt made to address this limitation can be found in Ouyang et al. (2014), who studied the optimal bus network configuration under spatially heterogeneous demand patterns.

Although the last reference investigated an irregular design of the bus network across a city, it still used a single trip length for all users in the network, similarly to all the previous studies. If, however, we look at the urban mobility patterns, we can see that such assumption is not realistic, since the trip lengths are not uniform. In addition, the authors considered only one mode (i.e. buses), without accounting for the collective effect of the analyzed topology in the global traffic performance. This has been relaxed in Yildirimoglu et al. (2016), who studied the impact of the design of transit system on the

available street capacities, modal split, and route choice, in Nikias et al. (2016), who studied the effects of bus network design on traffic, and in Loder et al. (2019a), who empirically evaluated the effects of bus operations on network's capacity and critical density. Nevertheless, none of these studies took into account the mode choice resulting from the complex bi-modal interactions and their significant effects on the traffic performance. A recent empirical study by Levinson (2012) found that the travel time is highly correlated with the topology of a road network. Hence, to maximize mobility in bi-modal urban systems, the impact of the bus network structure on the traffic performance should be carefully investigated.

To capture such interaction effects, we can use the recently proposed three-dimensional Macroscopic Fundamental Diagram (3D-MFD) (see e.g. Geroliminis et al., 2014; Loder et al., 2017; 2019b; Dakic et al., 2020a). The 3D-MFD allows to determine and model network-wide capacities and congestion levels, and describe city traffic macroscopically, while paying attention to multiple modes. As such, it offers new ways to analyze complex interactions between buses and cars (Dakic and Menendez, 2018), and has been mainly used to develop efficient perimeter control schemes (Ampountolas et al., 2017; He et al., 2019; Dakic et al., 2019) or bus dispatching policies (Dakic et al., 2021) for bi-modal urban networks. The potential to apply the 3D-MFD for the purpose of capturing traffic conditions for the bus network design problem has not been explored whatsoever.

2. Aims and contributions

In this paper, we aim to close the research gap and provide a general framework, capable of accounting for any trip length pattern, two types of user behavior, and the effects that the bus network structure might have on car traffic (and hence the passenger mode choice). For modeling different trip length patterns, we propose to use the trip length distribution as an intermediate level of abstraction. To quantify the operating speed for each mode, we apply the 3D-MFD. We then use these operating speeds to determine the mode choice at the trip length level. This way, we are able to determine the optimal bus network design in an analytical way, while accounting for possibly different types of user behavior and trip length patterns for the same level of passenger demand.

To achieve this goal, we extend the original problem defined by Daganzo (2010), relaxing several strong assumptions in the current state of the art: (i) absence of the complex modal interactions; (ii) absence of mode choice that depends on the design parameters; (iii) uniform trip length pattern; (iv) single type of user behavior; and (v) homogeneous design parameters including the headway and lane allocation layout across the network. By relaxing these important assumptions, we are able to provide insights on: (i) how to determine the optimal bus network configuration under mixed traffic; (ii) how the trip length patterns and user behavior affect the optimal bus network design; (iii) how the passenger demand and network topology affect the mode choice, hence the performance of the whole network, while taking into account multiple transport modes; (iv) how heterogeneous are the optimal design parameters across cardinal directions for a given level of passenger demand, trip length pattern, and user behavior; and (v) how important is to consider the complex modal interactions when determining the optimal bus network configuration. These insights would not have been possible without the proposed model.

Overall, the contributions of this research are sixfold. First, we account for spatially non-uniform network topology and distribution of the passenger demand across cardinal directions. Second, we use the distribution of the user trip lengths as an intermediate level of abstraction to determine the optimal design parameters. Such a level of abstraction allows not only to account for different trip length patterns (per and across cardinal directions) for the same level of passenger demand, but also to capture more accurately the network loading for all modes. Third, we consider the complex modal interactions and passenger mode choice, thus include the travel costs for both, the bus and the car mode, into the objective function. This way, we optimize the performance of the whole network while taking into account multiple transport modes. Fourth, we consider the influence of the passenger demand and network topology on the traffic performance, which, in turn, affects the passenger mode choice (i.e. demand). In other words, we use a full feedback loop to model all aforementioned dependencies. Fifth, we consider heterogeneous design parameters across cardinal directions, incorporating different lane allocations into the bus network design problem and accounting for their different influences on car traffic. Sixth, we investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network design and passenger mode choice. In addition to these contributions, we quantify the value of the proposed modeling framework.

For the readers convenience, Table 1 provides the list of the notation used in this paper.

The rest of the paper is organized as follows. Section 3 describes in detail the proposed methodological framework used to determine the optimal design parameters. In Section 4, we discuss the effects of demand intensity, user behavior, and trip length patterns on the optimal network configuration and passenger mode choice. Additionally, we quantify the value of considering complex modal interactions and heterogeneous bus network design across cardinal directions. Concluding remarks are given in Section 5.

3. Methodological framework

3.1. General network settings

We consider here a bi-modal, bi-directional urban network with an average trip generation rate Λ during the loading time of the peak period, and an average trip generation rate $\lambda < \Lambda$ during the off-peak periods and unloading time of the

Table 1
Nomenclature.

$\boldsymbol{D}$	set of decision variables
$\mathcal{M}$	set of modes, indexed by $m \in \mathcal{M}$
$\mathcal{A}$	set of possible lane allocations, indexed by $a \in \mathcal{A}$
$\mathcal{P}$	set of cardinal directions of travel, indexed by $p \in \mathcal{P}$
Λ	trip generation rate during the loading time of the peak hour
λ	trip generation rate during the unloading time of the peak hour and the off-peak periods
w	average walking speed
φ	expected number of transfers
α	average value of time
θ	parameter of the Logit model
δ	fixed penalty for transfers expressed in terms of an equivalent walking distance
ω	time lost per stop due to required door operations and deceleration/acceleration maneuver
ω'	time added per boarding/alighting passenger
Ω	average dwell time in the network
C	bus capacity
L_a	infrastructure length of lane allocation a
B	size of the bus fleet
V	total vehicular distance traveled by buses per hour of operation
$\pi_{L,a}/\pi_B/\pi_V$	unit infrastructure cost for lane allocation a/unit vehicle cost/unit distance cost
$Z_U/Z_O/Z$	user/operator/total system cost function
β	variable used to enumerate all cordons along E-W/N-S directions; $\beta \in [0, 1]$
$J_\beta(\ell_x)/J_\beta(\ell_y)$	event that bus users travel across vertical/horizontal cordon $\beta\phi_x/\beta\phi_y$ for a given ℓ_x/ℓ_y
ϕ_x/ϕ_y	network length along E-W/N-S directions
ψ_x/ψ_y	street spacing along E-W/N-S directions
s_x/s_y	stop spacing along E-W/N-S directions
l_x/l_y	line spacing along E-W/N-S directions
H_x/H_y	bus headway along E-W/N-S directions
N_x/N_y	number of corridors along E-W/N-S directions
$N_{x,b}/N_{y,b}$	number of bus lines along E-W/N-S directions
$z_{s,x}/z_{s,x}$	integer number of street spacings included in the stop spacing along E-W/N-S directions
$z_{l,x}/z_{l,x}$	integer number of stop spacings included in the line spacing along E-W/N-S directions
$\zeta_{x,m,a}/\zeta_{y,m,a}$	fraction of corridors with lane allocation a on which mode m operates along E-W/N-S directions
ℓ_x/ℓ_y	trip length realization along E-W/N-S directions
$f(\ell_x, \ell_y)$	joint probability density function of the user trip lengths
A_m	access (including the egress) time for mode m
W_m	waiting time for mode m
u_m	free flow speed of mode m
$\Pr(m)$	probability of choosing mode m
$\Pr(m\|\ell_x, \ell_y)$	probability of choosing mode m for a given combination of ℓ_x and ℓ_y
$TTT_{m\|\ell_x,\ell_y}$	total time traveled for mode m, for a given combination of ℓ_x and ℓ_y
$IVTT_{m\|\ell_x,\ell_y}$	in-vehicle time traveled for mode m, for a given combination of ℓ_x and ℓ_y
O_p	maximum bus occupancy in direction p
Q_p	maximum number of on-board passengers in direction p
η_p	number of lanes along a corridor in direction p
$v_{p,m}$	expected speed of mode m in direction p
$v_{p,m,a}$	operating speed of mode m for lane allocation a in direction p
$\tau_{p,c,a}$	experienced car delay along a corridor with lane allocation a in direction p
$\Lambda_{p,c,a}$	total generated car demand along a corridor with allocation a in direction p during the loading time of the peak period
$\mu_{p,c,a}$	maximum car outflow for lane allocation a in direction p given by the bi-modal MFD
$v^*_{p,c,a}$	operating car speed for lane allocation a in direction p corresponding to the maximum outflow
$P^*_{p,c,a}$	total car travel production for lane allocation a in direction p corresponding to the maximum outflow
$n^*_{p,c,a}$	total car accumulation for lane allocation a in direction p corresponding to the maximum outflow

peak period. Notice that, unlike the previous studies on the bus network design problem, we do not consider a mean value for the passenger demand over the full peak period, but distinguish between the loading and unloading phases. This is because we are interested in analyzing how congestion emerges across the network, allowing more realistic modeling of traffic conditions. That being said, the cost-terms (defined in the following subsection) are time-independent, i.e. they are estimated on an hourly base during the peak period and are given as an average of the loading and unloading phase. Also note that in this study we look at the bus network design problem under the steady state, when the traffic conditions are undesaturated or saturated, with no disruptions causing the system to enter a very congested state, i.e. there are no spillover effects. This is common in design applications, where one assumes that the network performs in an (under)saturated state (see e.g. Daganzo, 2010; Estrada et al., 2011; Nourbakhsh and Ouyang, 2012; Ouyang et al., 2014; Badia et al., 2014; Chen et al., 2015; Badia et al., 2016).

Let $\mathcal{M} = \{b, c\}$ be the set of modes, indexed by $m \in \mathcal{M}$, such that b denotes the bus and c denotes the car mode. Index cardinal directions of travel by $p \in \mathcal{P} = \{eb, wb, nb, sb\}$, where eb, wb, nb, and sb stand for the eastbound, westbound, northbound, and southbound directions, respectively. Similarly to previous studies, we assume that the network has a grid-like

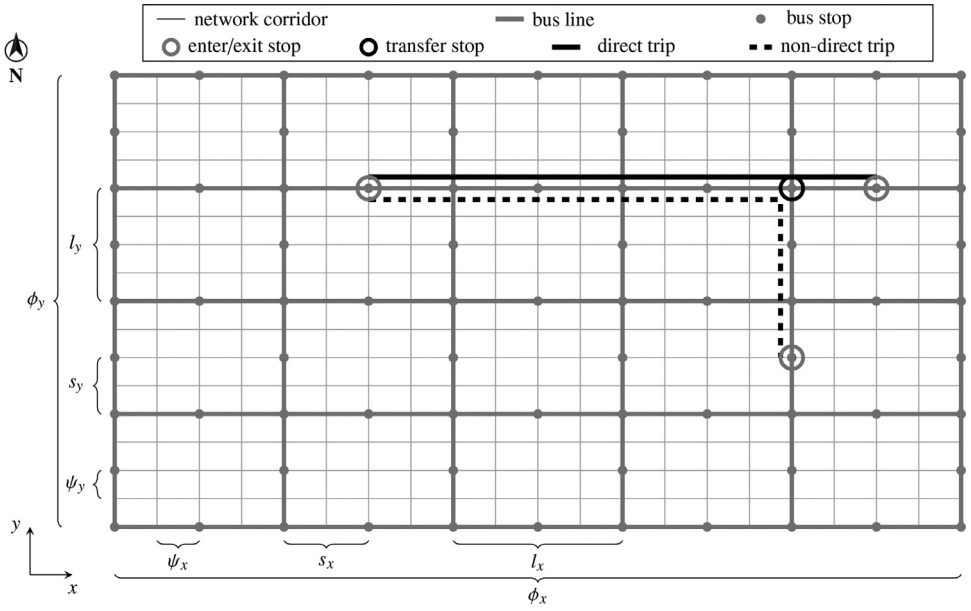

Fig. 1. City structure and the corresponding bus network.

structure. The reason for focusing on a grid-like configuration is because it is the most common network structure used in the literature, not only for the bus network design problem (see e.g. Daganzo, 2010; Estrada et al., 2011; Nourbakhsh and Ouyang, 2012; Ouyang et al., 2014; Chen et al., 2015; Badia et al., 2016; Amirgholy et al., 2017; Fan et al., 2018; Saade et al., 2018a), but also for many traffic-related problems (see e.g. Daganzo et al., 2011; Gayah and Daganzo, 2011; Zheng and Geroliminis, 2013; Gayah et al., 2014; Leclercq et al., 2014; Mühlich et al., 2015; Ortigosa et al., 2015; Girault et al., 2016; Ortigosa et al., 2017; 2019; Ambühl et al., 2019; Paipuri and Leclercq, 2020; Roca-Riu et al., 2020). To make a grid representation as realistic as possible, we assume a rectangular shape with sides ϕ_x and ϕ_y, as depicted in Fig. 1. Such a rectangular form is inspired by many cities (e.g. Barcelona, Manhattan, Buenos Aires, Oslo, Helsinki, Miami, and Washington D.C) that are elongated in shape (Estrada et al., 2011). The spacings between the streets along E-W (including *eb* and *wb*) and N-S (including *nb* and *sb*) directions are ψ_x and ψ_y, respectively. Atop this street structure is the bus network. The bus network consists of two types of bus lines that cross each other but do not overlap: (i) horizontal lines, with stop spacing s_x and line spacing l_y; and (ii) vertical lines, with stop spacing s_y and line spacing l_x (see Fig. 1).

Note that, although we focus on a grid bus network layout (as in e.g. Ouyang et al., 2014; Amirgholy et al., 2017; Saade et al., 2018b; Fan et al., 2018), we could still apply the proposed framework for combined network structures such as the hybrid one. This could be done by determining the grid bus network design in the city center using the proposed approach, and then distributing the bus lines across the peripheral part in a radial fashion (see Daganzo (2010) and Estrada et al. (2011) for more details). For radial or ring street layouts, however, we would need to extend our formulations. The formulations presented here can be considered as the first building block towards the design of public transport networks that consider for the first time the interactions with the traffic system, mode choice behavior, and heterogeneous demand patterns and types of user behavior, as elaborated further in this paper.

Following the approach by Estrada et al. (2011) and Fan et al. (2018), and for simplification purposes, we assume that the stop spacing includes an integer number of street segments, such that:

$$s_x = z_{s,x}\psi_x, \quad z_{s,x} \in \mathbb{N}, \tag{1a}$$

$$s_y = z_{s,y}\psi_y, \quad z_{s,y} \in \mathbb{N}. \tag{1b}$$

Likewise, the line spacing includes an integer number of stop spacings to facilitate transfers between the E-W and the N-S directions, i.e.:

$$l_x = z_{l,x}s_x = z_{l,x}z_{s,x}\psi_x, \quad z_{l,x}, z_{s,x} \in \mathbb{N}, \tag{2a}$$

$$l_y = z_{l,y}s_y = z_{l,y}z_{s,y}\psi_y, \quad z_{l,y}, z_{s,y} \in \mathbb{N}. \tag{2b}$$

Let ℓ_x and ℓ_y be the user trip lengths along the E-W and N-S directions, respectively. Denote by $\{f(\ell_x, \ell_y) : \ell_x \in [-\phi_x, \phi_x], \ell_y \in [-\phi_y, \phi_y]\}$ the probability density function of the user trip lengths, such that: $\ell_x \geq 0$ for *eb*; $\ell_x < 0$ for *wb*; $\ell_y \geq 0$ for *nb*; and $\ell_y < 0$ for *sb*. This distribution can be estimated by enumerating and grouping all OD pairs with the same ℓ_x and ℓ_y, using taxi GPS data (Yang et al., 2018), using mobile phone traces (Gonzalez et al., 2008; Paipuri et al., 2020), or

sampling other types of vehicle records (Batista et al., 2020). We consider the estimation of $f(\ell_x, \ell_y)$ to be out of the scope for this study, thus assume that $f(\ell_x, \ell_y)$ is exogenously given.

It should be noted here that by using $f(\ell_x, \ell_y)$ as an intermediate level of abstraction we lose the information about the locations of the trip initialization, which can affect the estimation of the maximum bus occupancy. That being said, the proposed framework is valid for all distributions of the user origins that are slowly-varying in space. This level of abstraction, however, allows to achieve not only a more realistic modeling of urban mobility patterns compared to the uniform distribution of user origins and destinations that comes with a unique trip distance, but also the analytical tractability of the problem. Nevertheless, it is worth noting that the proposed framework could also be applied for spatially-heterogeneous demand distributions. This could be done by clustering the passenger demand into different levels (e.g. *level 1* and *level 2*). Then, using the proposed approach, we would determine the optimal design parameters for *level 1* passenger demand. Afterwards, following the approach by Ouyang et al. (2014), we would increase the bus frequency in regions with higher demand densities (i.e. with *level 2* demand) by adding local bus routes parallel to the main ones (designed for *level 1* demand) in those regions (see Ouyang et al. (2014) for more details).

Also note that $f(\ell_x, \ell_y)$ has no time dependency. This is reasonable, as we design our network according to the demand in the peak period, during which we consider the average traffic conditions for the loading and unloading phase, hence the trip length distribution does not vary in time. This implies that for different peak periods (i.e. morning and afternoon) we can use different trip length distributions and determine the corresponding optimal bus network configurations. The final bus network design can then be found as the best out of the two.

In this study, we consider two types of user behavior: users adjust the walking distance at the origin and the destination to minimize the number of transfers while respecting a maximum threshold for the walking distance (Type 1); and users adjust the number of transfers to minimize the walking distance at the origin and the destination (Type 2). In each direction of travel, the bus demand is served by buses operating with a constant headway (H_x along E-W; H_y along N-S) and identical passenger capacity C, such that $C_x = C_y = C$ (i.e. all N-S lines have the same headway which is not necessarily the same as that of the E-W lines; and all buses are the same size). Furthermore, there are three types of corridors considered in this study: (i) corridors with car-only lanes; (ii) corridors with a dedicated bus lane; and (iii) corridors with a mixed lane. The share of the number of corridors of each type within the total number of corridors in the network directly affects the level of service provided by each individual mode, thus the passenger mode choice.

Let $\mathcal{A}_m$ be the set of types of corridors, indexed by $a \in \mathcal{A}_m$, along which mode m can operate. Assuming that cars can drive along all types of corridors in the network (i.e. the number of lanes along a corridor in any given direction p is $\eta_p \geq 2$), we set $\mathcal{A}_b = \{db, mb\}$ and $\mathcal{A}_c = \{co, db, mb\}$, where *co*, *db*, and *mb* stand for corridors with car-only lanes, corridors with a dedicated bus lane, and corridors with a mixed lane, respectively. Denote by $N_{x,b}$ the number of E-W corridors with a bus line and by $N_{y,b}$ the number of N-S corridors with a bus line, such that:

$$N_{x,b} = \sum_{a \in \mathcal{A}_b} N_{x,b,a} = \lfloor \phi_y / l_y \rfloor + 1, \tag{3a}$$

$$N_{y,b} = \sum_{a \in \mathcal{A}_b} N_{y,b,a} = \lfloor \phi_x / l_x \rfloor + 1, \tag{3b}$$

where $\lfloor \cdot \rfloor$ indicates the floor function; $N_{x,b,a}$ and $N_{y,b,a}$ represent the total number of corridors with a bus line operating with lane allocation $a \in \mathcal{A}_b$ along E-W and N-S directions, respectively. Note that the fraction of dedicated/mixed bus lines actually represents the fraction of bus lines with a dedicated/mixed bus lane. Since there is at most a single bus line per corridor, this fraction can also be interpreted as the fraction of bus lines running on dedicated/mixed bus lanes in a given cardinal direction. Then, assuming that the lane allocation remains the same along the whole corridor and for both travel directions (*eb* and *wb* along E-W; *nb* and *sb* along N-S), the fraction of bus lines with each lane allocation type along E-W and N-S directions can simply be obtained as:

$$\zeta_{x,b,a} = N_{x,b,a} / N_{x,b}, \quad a \in \mathcal{A}_b \tag{4a}$$

$$\zeta_{x,b,a} = N_{y,b,a} / N_{y,b}, \quad a \in \mathcal{A}_b. \tag{4b}$$

Similarly, the fraction of corridors available for car traffic operating with lane allocation $a \in \mathcal{A}_c$ along E-W ($\zeta_{x,c,a}$) and N-S ($\zeta_{y,c,a}$), can be computed using Eq. 5, where N_x is the number of E-W corridors and N_y is the number of N-S corridors in the network (Eq. 6).

$$\zeta_{x,c,a} = \begin{cases} 1 - N_{x,b}/N_x, & \text{if } a = co, \\ \zeta_{x,b,a} N_{x,b}/N_x, & \text{if } a \in \{db, mb\}, \end{cases} \tag{5a}$$

$$\zeta_{y,c,a} = \begin{cases} 1 - N_{y,b}/N_y, & \text{if } a = co, \\ \zeta_{y,b,a} N_{y,b}/N_y, & \text{if } a \in \{db, mb\}, \end{cases} \tag{5b}$$

with

$$N_x = \phi_y/\psi_y + 1, \tag{6a}$$

$$N_y = \phi_x/\psi_x + 1. \tag{6b}$$

Notice from Eqs. 4–5 that we do not specify which corridors have a dedicated bus lane. Instead, we stick to the fraction of dedicated bus lines along E-W and N-S directions as the global parameters for our design problem.

3.2. Mathematical formulation of the objective function

Here we formulate the objective function used to determine the optimal design parameters. The decision variables include stop spacings ($z_{s,x}$ and $z_{s,y}$), line spacings ($z_{l,x}$ and $z_{l,y}$), headways (H_x and H_y), and fractions of dedicated bus lines ($\zeta_{x,b,db}$ and $\zeta_{y,b,db}$) along E-W and N-S directions. They are denoted as $\boldsymbol{D} = [z_{s,x}, z_{s,y}, z_{l,x}, z_{l,y}, H_x, H_y, \zeta_{x,b,db}, \zeta_{y,b,db}]$. The objective function is given by Eq. 7 and, similarly to the previous studies, it consists of two components: agency cost $Z_O(\boldsymbol{D})$; and user cost $Z_U(\boldsymbol{D})$.

$$\min_{\boldsymbol{D}} Z(\boldsymbol{D}) = Z_O(\boldsymbol{D}) + Z_U(\boldsymbol{D}), \tag{7}$$

$$\text{s.t. } H_x, H_y \geq H_{\min}, \tag{8}$$

$$z_{s,x}, z_{s,y}, z_{l,x}, z_{l,y} \in \mathbb{N}, \tag{9}$$

$$0 \leq \zeta_{x,b,db} \leq 1, \;\; 0 \leq \zeta_{y,b,db} \leq 1, \tag{10}$$

$$O_p \leq C, \;\; \forall p \in \mathcal{P}. \tag{11}$$

where O_p is the maximum bus occupancy in direction p.

Note that all decision variables need to be strictly positive for physical reasons. In addition, the headway in each cardinal direction should not be below the predefined threshold $H_{\min}$ (Eq. 8). The stop spacings should include an integer number of street segments, whereas the line spacings should consist of an integer number of stop spacings (Eq. 9). Moreover, the number of dedicated bus lines cannot exceed the number of bus corridors (Eq. 10). Finally, the last constraint (Eq. 11) indicates that the maximum bus occupancy in each cardinal direction p needs to be lower than the bus passenger capacity C.

It is also worth noting that there is no need to incorporate constraints related to the connectivity or continuity of bus lanes, given that the bus lines operate along the entire corridor's length in all cardinal directions. In other words, both the connectivity and continuity of bus lanes are ensured by the general problem setup. They are connected to the level of service by explicitly considering and incorporating all parameters of the bus system in the operator and user cost function, as elaborated further below. Before we define the cost components, let us first introduce the assumptions used for the methodological framework, in terms of the lane allocation along a corridor, the waiting time of bus passengers, the vehicle dynamics, and the delay time experienced during the peak hour.

Assumption 1 (Lane allocation along a corridor)**.** We assume that the lane allocation remains the same along the whole corridor and for both travel directions.

Remark 1. Assumption 1 is made to prevent possible disruptions to the bus system induced by varying lane allocation, guaranteeing thereby the same operating conditions in both travel directions (He et al., 2018). Recall that this is a design problem, thus it would not be reasonable from a design point of view to disrupt operations by changing the lane allocation along a given arterial.

Assumption 2 (Waiting time of bus passengers)**.** We assume that bus users arrive at the stop independently of the schedule, thus they wait on average half of the bus headway.

Remark 2. Assumption 2 is reasonable for relatively short headways that are considered in this study (as in Daganzo, 2010; Estrada et al., 2011; Ouyang et al., 2014; Badia et al., 2016).

Assumption 3 (Vehicle dynamics)**.** We assume that, along corridors with mixed lane, buses drive at the speed of cars while cruising, i.e. both modes have the same speed either in the free-flow or congested traffic conditions.

Remark 3. Assumption 3 has been used in previous studies (see e.g. Daganzo, 2010; Estrada et al., 2011; Zheng and Geroliminis, 2013; Badia et al., 2016; Zhang and Liu, 2019) and was empirically confirmed (Dakic and Menendez, 2018). Nevertheless, it is only made for simplification purposes, given that the method used to compute the operating speed for each mode (Dakic et al., 2020a) allows to treat buses as moving bottlenecks while cruising. More details are given in Section 3.4.

Assumption 4 (Delays experienced during the peak hour)**.** We assume that the delays experienced during the peak hour in case the demand exceeds the maximum outflow can be approximated using a single bottleneck model.

Remark 4. Assumption 4 also used in Amirgholy et al. (2017) is made because we look at the bus network design problem under the steady state, when the traffic conditions are undersaturated or saturated, and there are no disruptions causing the network capacity to be dynamically reduced due to some internal or external factors. In other words, since the oversaturated conditions should not be solved by the design, but a proper control mechanism, we do not consider complex congestion spreading during very congested states. More details are given in Section 3.4.

The agency cost metrics include the infrastructure length of each lane allocation type $\{L_a(\boldsymbol{D}) : a \in \mathcal{A}_b\}$, the size of the bus fleet $B(\boldsymbol{D})$, and the total vehicular distance traveled by buses per hour of operation $V(\boldsymbol{D})$ (Eq. 12). The associated parameters $\{\pi_{L,a} : a \in \mathcal{A}_b\}$, π_B, and π_V are the corresponding unit costs. To put all the cost components in the objective function (Eq. 7) under the same unit (i.e. hours per passenger), we convert agency money into an equivalent user riding time with parameter α denoting an average value of time, and divide it by the total number of bus passengers, $\Pr(b)\lambda$. Note that we use lower trip generation rate value (λ) in Eq. 12 as is gives us the maximum bound for the operator cost, which corresponds to the worst-case scenario (see Daganzo (2010) for more details).

$$Z_O(\boldsymbol{D}) = (\pi_{L,db} L_{db}(\boldsymbol{D}) + \pi_{L,mb} L_{mb}(\boldsymbol{D}) + \pi_B B(\boldsymbol{D}) + \pi_V V(\boldsymbol{D}))/(\alpha \Pr(b)\lambda), \tag{12}$$

with

$$\Pr(m) = \int_{-\phi_x}^{\phi_x} \int_{-\phi_y}^{\phi_y} \Pr(m|\ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x, \quad m \in \mathcal{M}, \tag{13}$$

where $\Pr(m|\ell_x, \ell_y)$ stands for the probability of choosing mode m for a given combination of ℓ_x and ℓ_y. This probability can be computed by applying the Logit model (Bhat and Guo, 2004), as given by Eq. 14, where θ is the parameter of the Logit model, and $TTT(\boldsymbol{D})_{m|\ell_x,\ell_y}$ denotes the total time traveled for a given mode m and combination of ℓ_x and ℓ_y.

$$\Pr(m|\ell_x, \ell_y) = \frac{\exp(-\theta \cdot TTT(\boldsymbol{D})_{m|\ell_x,\ell_y})}{\sum_{m' \in \mathcal{M}} \exp(-\theta \cdot TTT(\boldsymbol{D})_{m'|\ell_x,\ell_y})}, \quad m \in \mathcal{M}. \tag{14}$$

On the other hand, the user cost metric of interest is the expected total time traveled, computed across both modes and the entire set of the user trip lengths (Eq. 15).

$$Z_U(\boldsymbol{D}) = \int_{-\phi_x}^{\phi_x} \int_{-\phi_y}^{\phi_y} TTT(\boldsymbol{D})_{m|\ell_x,\ell_y} \Pr(m|\ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x. \tag{15}$$

In the following, we derive the components of the operator cost function. We start by formulating the infrastructure length (Eq. 16). This variable can simply be obtained by multiplying the length associated with one bus line by the total number of bus lines with a given lane allocation along E-W and N-S directions.

$$L_a(\boldsymbol{D}) = \begin{cases} \zeta_{x,b,db} N_{x,b} \phi_x + \zeta_{y,b,db} N_{y,b} \phi_y, & \text{if } a = db, \\ (1 - \zeta_{x,b,db}) N_{x,b} \phi_x + (1 - \zeta_{y,b,db}) N_{y,b} \phi_y, & \text{if } a = mb. \end{cases} \tag{16}$$

The bus fleet consists of buses operating in all cardinal directions (Eq. 17). In each direction of travel, the number of required buses can be approximated by multiplying the number of buses required for one bus line by the total number of bus lines.

$$B(\boldsymbol{D}) = \sum_{p \in \{eb,wb\}} N_{x,b} \phi_x / (H_x v_{p,b}) + \sum_{p \in \{nb,sb\}} N_{y,b} \phi_y / (H_y v_{p,b}), \tag{17}$$

where $v_{p,b}$ stands for the expected bus speed (including the dwell and acceleration/deceleration time) in direction p. This expected bus speed can be determined using Eq. 18, where $v_{p,m,a}$ denotes the operating speed of mode m for lane allocation a in direction p.

$$v_{p,m} = \begin{cases} \sum_{a \in \mathcal{A}_m} v_{p,m,a} \zeta_{x,m,a}, & \text{if } p \in \{eb, wb\}, \\ \sum_{a \in \mathcal{A}_m} v_{p,m,a} \zeta_{y,m,a}, & \text{if } p \in \{nb, sb\}, \end{cases} \quad m \in \mathcal{M} \tag{18}$$

Note that the operating speeds $v_{p,m,a}$ are not predefined, but depend on the network loading and the topology in a given cardinal direction. This will be further elaborated in Section 3.4. Following the approach by Daganzo (2010), we formulate the bus operating speed as follows:

$$v_{p,b,a} = \begin{cases} (1/u_b + \omega/s_x + \Omega)^{-1}, & \text{if } p \in \{eb, wb\}, \ a = db \\ (1/u_b + \omega/s_y + \Omega)^{-1}, & \text{if } p \in \{nb, sb\}, \ a = db, \\ (1/v_{p,c,mb} + \omega/s_x + \Omega)^{-1}, & \text{if } p \in \{eb, wb\}, \ a = mb, \\ (1/v_{p,c,mb} + \omega/s_y + \Omega)^{-1}, & \text{if } p \in \{nb, sb\}, \ a = mb, \end{cases} \tag{19}$$

where u_b is the bus free-flow speed; $v_{p,c,mb}$ denotes the operating car speed along a mixed lane in direction p; ω represents the time lost per stop due to required door operations and deceleration/acceleration maneuvers; Ω stands for the average dwell time in the network. To estimate Ω, we first need to define the last agency cost metric, the total vehicular distance

traveled by buses per hour of operation $V(\boldsymbol{D})$. Following the approach by Daganzo (2010), we define this distance as the ratio of the length of the network covered by bus lines and the corresponding bus headway across all cardinal directions. It can also be defined as the product of the total number of buses operating in all cardinal directions and the corresponding expected bus speeds (Eq. 20). Note that, since the headway for E-W/N-S bus lines is the same in both travel directions (east to west and west to east / north to south and south to north), we multiply the total vehicular distance traveled along E-W/N-S direction by 2.

$$V(\boldsymbol{D}) = 2(N_{x,b}\phi_x/H_x + N_{y,b}\phi_y/H_y). \tag{20}$$

Then, the average dwell time in the network can be obtained by multiplying the time added per boarding/alighting passenger (ω') by the average number of passengers a vehicle collects per unit distance traveled:

$$\Omega = \omega' \Pr(b)\Lambda(1+\varphi(\boldsymbol{D}))/V(\boldsymbol{D}). \tag{21}$$

where $\varphi(\boldsymbol{D})$ is the expected number of transfers. According to Daganzo (2010), the average number of passengers a vehicle collects per unit distance traveled can be approximated as the ratio of the number of boarding passengers generated per hour during the peak period ($\Pr(b)\cdot\Lambda\cdot(1+\varphi)$) and the total vehicular distance traveled by buses per hour of operation. To get the maximum bound for the user cost function, we use higher trip generation rate value (Λ) in Eq. 21, similarly to Daganzo (2010) and Estrada et al. (2011).

Finally, the last term in the objective function is the total time traveled for a given mode m and combination of the user trip lengths. It consists of the expected access (including the egress) time $A_m(\boldsymbol{D})$, the expected waiting time $W_m(\boldsymbol{D})$, the in-vehicle time traveled $IVTT(\boldsymbol{D})_{m|\ell_x,\ell_y}$ for a given ℓ_x and ℓ_y, and some additional time components (Eq. 22). For the bus mode, these additional components include the transferring time, whereas for the car mode, the additional time traveled is related to the equivalent mileage costs of an auto trip (as in Daganzo, 2010).[1]

$$TTT(\boldsymbol{D})_{m|\ell_x,\ell_y} = A_m(\boldsymbol{D}) + W_m(\boldsymbol{D}) + IVTT(\boldsymbol{D})_{m|\ell_x,\ell_y} + \mathbb{1}_{\{m=b\}}\varphi(\boldsymbol{D})\delta/w + \mathbb{1}_{\{m=c\}}(|\ell_x|+|\ell_y|)\pi_D/\alpha, \quad m\in\mathcal{M}, \tag{22}$$

with

$$IVTT(\boldsymbol{D})_{m|\ell_x,\ell_y} = |\ell_x|/v_{\xi(\ell_x),m} + |\ell_y|/v_{\xi(\ell_y),m}, \quad m\in\mathcal{M}, \tag{23}$$

where δ is a fixed penalty for transfers expressed in terms of an equivalent walking distance; w denotes the average walking speed (accounting for the delays that users encounter when crossing streets); π_D represents the auto cost per unit distance; $\mathbb{1}_{\{condition\}}$ is an indicator function that return the value of 1 if *condition* is satisfied; $\xi(\cdot)$ stands for a function that maps each trip length component (ℓ_x and ℓ_y) to a cardinal direction in which it is traversed, such that: $\xi(\ell_x\geq 0)=eb$; $\xi(\ell_x<0)=wb$; $\xi(\ell_y\geq 0)=nb$; and $\xi(\ell_y<0)=sb$.

The following equations approximate the network-level components of the user cost function for the bus mode (Eq. 22), including the access, the waiting, and the transferring time, for each type of user behavior. For Type 1, Eqs. 24–26 represent an extension of those in Estrada et al. (2011) for more general bus network configurations. We say an extension, because the formulations proposed by Estrada et al. (2011) are only applicable for networks with the same stop spacing and bus headway along E-W and N-S directions (i.e. $s_x=s_y$ and $H_x=H_y$), which may not be the case in this study (i.e. design parameters may vary across cardinal directions).

Type 1. (adjusting the walking distance to minimize the number of transfer)

$$A_{1,b}(\boldsymbol{D}) = (l_x+s_y+l_y+s_x)/(4w). \tag{24}$$

$$\varphi_1(\boldsymbol{D}) = \Pr(\varphi>0) = 1-\Pr(\varphi=0) = 1-(l_y\phi_x+l_x\phi_y-l_xl_y)/(\phi_x\phi_y). \tag{25}$$

$$W_{1,b}(\boldsymbol{D}) = (1+\varphi_1(\boldsymbol{D}))\bar{H}/2. \tag{26}$$

where $\Pr(\varphi=0)$ and $\Pr(\varphi>0)$ are the probabilities for making zero and non-zero transfers, respectively; $\bar{H}=(N_{b,x}H_x+N_{b,y}H_y)/(N_{b,x}+N_{b,y})$ denotes the weighted (by the number of bus lines in the corresponding cardinal directions) bus headway in the network.

Type 2. (adjusting the number of transfer to minimize the walking distance)

$$A_{2,b}(\boldsymbol{D}) = \begin{cases} (6l_xl_y+6l_xs_x-2l_y^2-3l_ys_x+3l_ys_y)/(12l_xw), & \text{if } l_x\geq l_y,\\ (6l_xl_y+6l_ys_y-2l_x^2-3l_xs_y+3l_xs_x)/(12l_yw), & \text{if } l_x<l_y. \end{cases} \tag{27}$$

$$\varphi_2(\boldsymbol{D}) = \begin{cases} \Pr(\varphi>0)(4l_x^2l_y^2-2l_xl_y^3+l_y^4)/(2l_x^2l_y^2), & \text{if } l_x\geq l_y,\\ \Pr(\varphi>0)(4l_x^2l_y^2-2l_x^3l_y+l_x^4)/(2l_x^2l_y^2), & \text{if } l_x<l_y. \end{cases} \tag{28}$$

$$W_{2,b}(\boldsymbol{D}) = \begin{cases} \Pr(\varphi=0)\bar{H}/2+\Pr(\varphi>0)(H_x(8l_x^2l_y^2-4l_xl_y^3+l_y^4)+H_y(4l_x^2l_y^2+l_y^4))/(8l_x^2l_y^2), & \text{if } l_x\geq l_y,\\ \Pr(\varphi=0)\bar{H}/2+\Pr(\varphi>0)(H_x(4l_x^2l_y^2+l_x^4)+H_y(8l_x^2l_y^2-4l_x^3l_y+l_x^4))/(8l_x^2l_y^2), & \text{if } l_x<l_y. \end{cases} \tag{29}$$

[1] Alternatively, one can also incorporate the fare for using the bus mode and/or congestion tolls, parking, and garaging costs associated with the car mode.

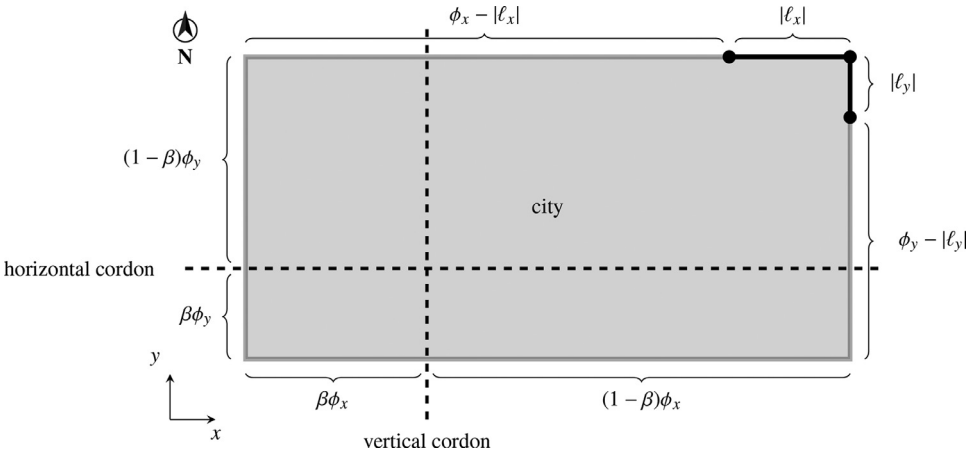

Fig. 2. A city and its vertical and horizontal cordons.

3.3. Derivation of the maximum bus occupancy

Now that we have derived all components of the objective function, we can compute the maximum bus occupancy O_p in each cardinal direction p. This variable is important to ensure that the bus service provided is enough to accommodate the passenger demand generated in the most restrictive links during the peak hour.

To compute O_p, we have to evaluate the maximum number of on-board passengers Q_p across all points in the corresponding cardinal direction. For this purpose, we use vertical and horizontal cordons denoted as $\beta \cdot \phi_x$ and $\beta \cdot \phi_y$, respectively, such that $\beta \in [0, 1]$ (see Fig. 2). Intuitively, vertical cordons are used for Q_{eb} and Q_{wb}, whereas the horizontal ones are used for Q_{nb} and Q_{sb}. Recall that due to the use of the trip length distribution as an intermediate level of abstraction, we do not know the locations of the trip initialization. Hence, for each cordon, we have to determine the probability that the bus users cross the cordon for a given ℓ_x (in case of a vertical cordon) or ℓ_y (in case of a horizontal cordon), in which case they are considered as part of the total number of on-board passengers in the corresponding cardinal direction. These probabilities are denoted as $\Pr(J_\beta(\ell_x))$ and $\Pr(J_\beta(\ell_y))$, where $J_\beta(\ell_x)$ is the event that the bus users travel across vertical cordon $\beta\phi_x$ for a given ℓ_x, and $J_\beta(\ell_y)$ is the event that the bus users travel across horizontal cordon $\beta\phi_y$ for a given ℓ_y. Given the assumption regarding the slowly-varying spatial distribution of the user origins, the aforementioned probabilities can be computed as:

$$\Pr(J_\beta(\ell_x)) = \begin{cases} |\ell_{x,b}|/(\phi_x - |\ell_{x,b}|), & \text{if } |\ell_{x,b}| \le \min\{\beta\phi_x, (1-\beta)\phi_x\}, \\ \min\{\beta\phi_x, (1-\beta)\phi_x\}/(\phi_x - |\ell_{x,b}|), & \text{if } \min\{\beta\phi_x, (1-\beta)\phi_x\} < |\ell_{x,b}| \le \max\{\beta\phi_x, (1-\beta)\phi_x\}, \\ 1, & \text{if } |\ell_{x,b}| > \max\{\beta\phi_x, (1-\beta)\phi_x\}, \end{cases} \tag{30a}$$

$$\Pr(J_\beta(\ell_y)) = \begin{cases} |\ell_{y,b}|/(\phi_y - |\ell_{y,b}|), & \text{if } |\ell_{y,b}| \le \min\{\beta\phi_y, (1-\beta)\phi_y\}, \\ \min\{\beta\phi_y, (1-\beta)\phi_y\}/(\phi_y - |\ell_{y,b}|), & \text{if } \min\{\beta\phi_y, (1-\beta)\phi_y\} < |\ell_{y,b}| \le \max\{\beta\phi_y, (1-\beta)\phi_y\}, \\ 1, & \text{if } |\ell_{y,b}| > \max\{\beta\phi_y, (1-\beta)\phi_y\}, \end{cases} \tag{30b}$$

where $\phi_x - |\ell_x|$ and $\phi_y - |\ell_y|$ represent the domains of valid user origins for the corresponding ℓ_x and ℓ_y (see Fig. 2).

Then, the maximum bus occupancy can simply be obtained by dividing the maximum number of on-board passengers across all bus lines and all cordons in a given cardinal direction by the total number of bus lines and the respective bus frequency (Eq. 31). In other words, by moving the cordons in a given cardinal direction, we enumerate all possible occupancy rates, based on which we find the maximum one for the corresponding direction of travel. Note that we use the higher trip generation rate in Eq. 32, i.e. the value of Λ observed during the loading time of the peak period, to get the maximum number of on-board passengers.

$$O_p = \begin{cases} Q_p H_x / N_{x,b}, & \text{if } p \in \{eb, wb\}, \\ Q_p H_y / N_{y,b}, & \text{if } p \in \{nb, sb\}, \end{cases} \tag{31}$$

with

$$Q_{eb} = \max_{\beta\in[0,1]} \left\{ \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \Pr(J_\beta(\ell_x)) \Lambda \Pr(b|\ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x \right\}, \tag{32a}$$

$$Q_{wb} = \max_{\beta\in[0,1]} \left\{ \int_{-\phi_x}^{0} \int_{-\phi_y}^{\phi_y} \Pr(J_\beta(\ell_x)) \Lambda \Pr(b|\ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x \right\}, \tag{32b}$$

$$Q_{nb} = \max_{\beta\in[0,1]} \left\{ \int_0^{\phi_y} \int_{-\phi_x}^{\phi_x} \Pr(J_\beta(\ell_y)) \Lambda \Pr(b|\ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_x \mathrm{d}\ell_y \right\}, \tag{32c}$$

$$Q_{sb} = \max_{\beta\in[0,1]} \left\{ \int_{-\phi_y}^{0} \int_{-\phi_x}^{\phi_x} \Pr(J_\beta(\ell_y)) \Lambda \Pr(b|\ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_x \mathrm{d}\ell_y \right\}. \tag{32d}$$

Notice from Eq. 31 the importance of providing different bus headways along E-W and N-S directions. In other words, a unique bus headway for the entire network might lead to an efficient bus service (measured by the ability to serve the total generated demand with the available bus capacity) in some directions. However, other directions might experience a lower level of service, i.e. significantly higher bus occupancy, due to its less favorable design parameters (e.g. less number of bus lines, lower fraction of dedicated bus lines, or longer stop spacing), constrained by the geometrical characteristics of the network. In other words, a unique bus headway, as it has been previously proposed in most of the literature (see e.g. Daganzo, 2010; Estrada et al., 2011; Ouyang et al., 2014; Amirgholy et al., 2017), might not lead to the optimal design parameters in all cardinal directions of travel. Therefore, by considering different bus headways, we balance the bus level of service provided in each direction, preventing such scenario.

To verify Eqs. 30–32, we show that the solution for the maximum bus occupancy in case of uniformly and independently distributed origins and destinations is the same to that of Daganzo (2010) (see the Appendix for more details).

3.4. Derivation of the operating car speed for each lane allocation type

As mentioned before, in the problem investigated in this study, there are three types of corridor in each cardinal direction of travel p for which the operating speed $v_{p,m,a}$ needs to be estimated: (i) corridors with car-only lanes; (ii) corridors with a dedicated bus lane; and (iii) corridors with a mixed lane. For corridors with a dedicated bus lane, the operating bus speed $v_{p,b,db}$ is not affected by the traffic conditions, i.e. interactions with the car mode (assuming that a dedicated bus lane operates with a transit signal priority, as in Wu and Guler, 2018). Hence, it can simply be modeled as the free-flow bus speed u_b (including the dwell and acceleration/deceleration time) (Eq. 19). Likewise, given Assumption 3 that buses drive at the speed of cars while cruising, the operating bus speed for corridors with a mixed lane, $v_{p,b,mb}$, can be modeled as the operating car speed along a corridor with a mixed lane in the corresponding direction of travel, $v_{p,c,mb}$ (including the dwell time and acceleration/deceleration time) (Eq. 19). As a result, for each cardinal direction and type of corridors, we only need to estimate the operating car speed. This operating car speed depends on the accumulation of both buses and cars.

One of the commonly used approaches to determine the operating speed for each mode at the network level, while capturing the modal interactions and the network topology, is the three dimensional macroscopic fundamental diagram (3D-MFD) (Geroliminis et al., 2014; Loder et al., 2017; 2019b; Dakic et al., 2020a; Paipuri and Leclercq, 2020). Note that by the modal interactions we refer to how cars and buses influence each other in terms of traffic dynamics. In particular, such interactions include the impact of bus operations (e.g. accumulation, headway, dwell times at bus stops along mixed lanes) on the speed of cars, as well as the impact of car accumulation on the operation of buses (e.g. reduced speed along mixed lanes).

In the classical framework, the 3D-MFD is assumed to be a function of the total production in the network with respect to the accumulation of both buses and cars. Such 3D-MFD is resourceful for practitioners to understand the impact of each mode in the network. Nevertheless, for modeling purposes, it is more appealing to segregate the 3D-MFD into two partial 3D-MFDs, one for each mode of transport (Paipuri and Leclercq, 2020). This segregation yields a more accurate estimation of the operating speed for each mode. Therefore, in this paper, we apply the partial car 3D-MFD to determine the operating car speed and ultimately the optimal design parameters. It is worth noting that we could also apply the partial bus 3D-MFD to determine the operating bus speed along corridors with a mixed lane. However, this would not yield a very accurate bus speed estimation for a given level of bus passenger demand, given that the partial bus 3D-MFD does not explicitly account for the dwell time that depends on the bus passenger demand level. For that reason, we compute the bus speed along corridors with a mixed lane using the operating car speed (determined from the partial car 3D-MFD) and the analytical expression for the dwell time as a function of the bus passenger demand (Eq. 21).

To estimate the partial car 3D-MFD, we combine bi-modal MFDs for the car mode using the framework proposed by Dakic et al. (2020a). Each bi-modal MFD represents a slice of the partial car 3D-MFD, covering a range of traffic conditions for a particular bus headway. Different traffic conditions are associated with different cruising bus speeds. Those cruising bus speeds define the temporal distribution of bus stop shortcuts in the variational theory graph. When modeling the temporal distribution of bus stop shortcuts, we keep either the bus accumulation or headway constant, depending on whether

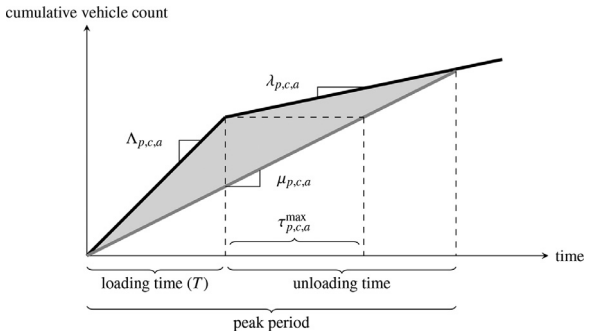

Fig. 3. Computing the average car delay experienced along a corridor with allocation a in direction p.

the operator operates with a constant headway or a constant number of buses (for a detailed overview, please refer to Dakic et al., 2020a). Thus, to change the operating regime of buses with the decision variables (i.e. bus headways H_x and H_y), we use different slices of the obtained partial car 3D-MFD, as long as the network topology (defined by the decision variables $\zeta_{x,b,db}$ and $\zeta_{y,b,db}$) remains the same. Once the network topology changes, we derive the corresponding partial car 3D-MFD by repeating the process mentioned above. This way, we are able to determine the effects that the design variables have on the interactions between the modes. We do so by quantifying the operating car speed (and hence the operating bus speed) for any given bus network configuration scenario (as defined by the decision variables).

Note, however, that most analytical studies for approximating an MFD (or the bi-modal MFD) using variational theory treat the entire network as a single, one dimensional corridor (Daganzo and Geroliminis, 2008). This may not lead to a very accurate bi-modal MFD approximation for networks whose topological (intersection and stop spacing) and operational characteristics (signal settings, lane allocation, and bus headway) vary significantly across cardinal directions, as it may be the case in the considered problem. Therefore, to more accurately quantify the performance of each mode, we estimate the 3D-MFD for each cardinal direction and corridor type, i.e. lane allocation layout (corridors with car only lanes, a dedicated bus lane, or a mixed lane), and then combine the results similar to Girault et al. (2016). It is important to note that combining directional (3D-)MFDs is reasonable for the considered bus network design problem, given that we examine the best case scenario when there are no disruptions causing the network capacity to be dynamically reduced due to some internal or external factors. In other words, the traffic control operates well-enough to remove local capacity drops. Consequently, the outflow in one direction can smoothly go in the other. This is a regular process when discussing the design problem. Once the 3D-MFD is estimated for each travel direction and type of corridor, we feed the derived 3D-MFD with the corresponding travel demand, as elaborated further below. This way, we implicitly take into account the interactions between cardinal directions. Interactions across the two cardinal directions include signal timing allocation (which we do take into account), demand distribution (which we also take into account), and spillovers (which are not relevant for a design problem, as discussed in Section 3.1).

To compute $\nu_{p,c,a}$, we first need to determine whether the total generated car demand $\Lambda_{p,c,a}$ (Eq. 33) along a corridor with allocation a in direction p during the loading time T of the peak period (Fig. 3) is smaller than the corresponding corridor capacity $\mu_{p,c,a} = P^*_{p,c,a}/\bar{\ell}_p$. We define the corridor capacity as the maximum outflow, i.e. ratio of the maximum travel production, $P^*_{p,c,a}$, and the average trip length for a given cardinal direction, $\bar{\ell}_p$ (Eq. 34). Notice from Eq. 33 that we assume that the car capacity for a corridor with a dedicated bus line is reduced by one lane compared to other lane allocation types.

$$\Lambda_{eb,c,a} = \int_0^{\phi_x}\int_{-\phi_y}^{\phi_y} (\Lambda/N_x(\eta_p - \mathbb{1}_{\{a=db\}}))\Pr(c|\ell_x,\ell_y)f(\ell_x,\ell_y)\mathrm{d}\ell_y\mathrm{d}\ell_x, \tag{33a}$$

$$\Lambda_{wb,c,a} = \int_{-\phi_x}^{0}\int_{-\phi_y}^{\phi_y} (\Lambda/N_x(\eta_p - \mathbb{1}_{\{a=db\}}))\Pr(c|\ell_x,\ell_y)f(\ell_x,\ell_y)\mathrm{d}\ell_y\mathrm{d}\ell_x, \tag{33b}$$

$$\Lambda_{nb,c,a} = \int_0^{\phi_y}\int_{-\phi_x}^{\phi_x} (\Lambda/N_y(\eta_p - \mathbb{1}_{\{a=db\}}))\Pr(c|\ell_x,\ell_y)f(\ell_x,\ell_y)\mathrm{d}\ell_x\mathrm{d}\ell_y, \tag{33c}$$

$$\Lambda_{sb,c,a} = \int_{-\phi_y}^{0}\int_{-\phi_x}^{\phi_x} (\Lambda/N_y(\eta_p - \mathbb{1}_{\{a=db\}}))\Pr(c|\ell_x,\ell_y)f(\ell_x,\ell_y)\mathrm{d}\ell_x\mathrm{d}\ell_y. \tag{33d}$$

$$\bar{\ell}_{eb} = \frac{\int_0^{\phi_x}\int_{-\phi_y}^{\phi_y} |\ell_x| f(\ell_x,\ell_y)\mathrm{d}\ell_y\mathrm{d}\ell_x}{\int_0^{\phi_x}\int_{-\phi_y}^{\phi_y} f(\ell_x,\ell_y)\mathrm{d}\ell_y\mathrm{d}\ell_x}, \tag{34a}$$

$$\bar{\ell}_{wb} = \frac{\int_{-\phi_x}^{0}\int_{-\phi_y}^{\phi_y} |\ell_x| f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x}{\int_{-\phi_x}^{0}\int_{-\phi_y}^{\phi_y} f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x}, \tag{34b}$$

$$\bar{\ell}_{nb} = \frac{\int_{0}^{\phi_y}\int_{-\phi_x}^{\phi_x} |\ell_y| f(\ell_x, \ell_y) \mathrm{d}\ell_x \mathrm{d}\ell_y}{\int_{0}^{\phi_y}\int_{-\phi_x}^{\phi_x} f(\ell_x, \ell_y) \mathrm{d}\ell_x \mathrm{d}\ell_y}, \tag{34c}$$

$$\bar{\ell}_{sb} = \frac{\int_{-\phi_y}^{0}\int_{-\phi_x}^{\phi_x} |\ell_y| f(\ell_x, \ell_y) \mathrm{d}\ell_x \mathrm{d}\ell_y}{\int_{-\phi_y}^{0}\int_{-\phi_x}^{\phi_x} f(\ell_x, \ell_y) \mathrm{d}\ell_x \mathrm{d}\ell_y}. \tag{34d}$$

Depending on the ratio between $\Lambda_{p,c,a}$ and $\mu_{p,c,a}$, we can distinguish the following two cases: $\Lambda_{p,c,a} \leq \mu_{p,c,a}$; and $\Lambda_{p,c,a} > \mu_{p,c,a}$. The first case implies uncongested traffic conditions, in which the corridor capacity is sufficient to serve the total generated demand, hence no delays due to congestion are experienced in the system. Consequently, there is a direct match between the total generated demand and the outflow along a corridor. In the second case, however, such a direct match between the demand and the outflow can no longer be established, as the outflow might be limited to the corridor capacity during the time when congestion is active. Therefore, to tackle this problem and given Assumption 4, we use a simple bottleneck model that considers the entire corridor as a single bottleneck with a mean capacity (see Amirgholy et al. (2017) for more details). Note that this model is applied to account for possible delays during the peak hour in case the demand exceeds the maximum outflow. It is used because we look at the bus network design problem under the steady state, when the traffic conditions are (under)saturated and there are no disruptions causing the network capacity to be dynamically reduced due to some internal or external factors, as we stated before. Oversaturated conditions should not be solved by the design, but a proper control mechanism (e.g. He et al., 2018). Hence, we do not consider complex congestion spreading during very congested states. In the following, we summarize how the operating car speed is computed for each case separately.

Case 1. ($\Lambda_{p,c,a} \leq \mu_{p,c,a}$): the network operates in the uncongested regime. The operating car speed can be determined by finding the point on the bi-modal MFD in which the total travel production is equivalent to that obtained by applying the queuing formula of Little (1961):

$$P_{p,c,a} = \Lambda_{p,c,a}\bar{\ell}_p, \quad p \in \mathcal{A}_c, \ p \in \mathcal{P}. \tag{35}$$

Case 2. ($\Lambda_{p,c,a} > \mu_{p,c,a}$): the network operates at capacity. Due to congestion, cars experience certain delay $\tau_{p,c,a}$, which needs to be taken into account when computing $v_{p,c,a}$. Following the approach by Amirgholy et al. (2017), we can compute this delay as half of the maximum car delay $\tau_{p,c,a}^{\max}$ experienced during the peak hour, i.e. $\tau_{p,c,a} = \tau_{p,c,a}^{\max}/2 = (T(\Lambda_{p,c,a} - \mu_{p,c,a})/\mu_{p,c,a})/2$ (see Fig. 3). The operating car speed is then given by Eq. 36, where $v_{p,c,a}^{*}$ denotes the operating car speed corresponding to the maximum outflow along a corridor with allocation a in direction p.

$$v_{p,c,a} = (1/v_{p,c,a}^{*} + \tau_{p,c,a})^{-1}, \quad p \in \mathcal{P}, \ a \in \mathcal{A}_c \tag{36}$$

4. Numerical results

4.1. Solution algorithm

For the purpose of computing the optimal design parameters, we follow the approach of previous studies and use a brute-force search technique (see e.g. Daganzo, 2010; Estrada et al., 2011; Badia et al., 2014; 2016). Alternatively, other optimization techniques for solving the problem could also be explored, which may give the solution faster. However, this is out of the scope for this paper. To speed up the searching process over the feasible region, we perform a Monte Carlo sampling for a given trip length pattern. Notice that more elaborated sampling techniques could potentially be used to further increase the computational speed (Ge and Menendez, 2017). The result is the set of trip length values along E-W ($\mathcal{L}_x$) and N-S ($\mathcal{L}_y$) directions. Then, we find the network-level parameters of the bus user cost function according to the selected type of user behavior (i.e. minimizing the number of transfers or the walking distance). To compute the total travel cost for each mode, we need to find the operating speeds. These operating speeds depend on the mode choice that is, on the other hand, a function of the utility cost (i.e. travel time) for each mode. The utility costs are determined based on the bi-modal MFD derived for a given bus network configuration scenario (defined by the decision variables). In other words, the total travel costs are computed by solving a fixed point problem, as shown by the flow chart in Fig. 4. This flow chart summarizes the methodological steps.

Note that, rather than using a conventional convergence algorithm to solve the fixed point problem, we discretize (by 1%) potential mode choice values, and for each value we compute its absolute difference to the mode choice value given by the Logit model Eqs. 13–(14). Then, using a brute-force approach, we find the value of the mode choice that has the minimum absolute difference compared to one obtained from the Logit model. This minimum absolute difference found is always less than 1%, given that we discretize our mode choice values by 1%. Once the fixed point problem is solved, we compute the operator and user costs, hence the total system cost.

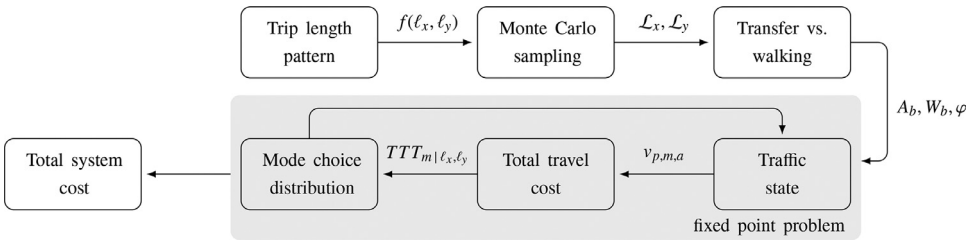

Fig. 4. Schematic representation of the solution algorithm.

Table 2
Experimental settings.

Input parameter	Variable	Units	Value
Network length	ϕ_x/ϕ_y	km	10/4.95
Street spacings	ψ_x/ψ_y	km	0.25/0.15
Equivalent penalty distance per transfer	δ	km	0.03
Trip generation rate during the peak hour	Λ	pax/hr	75,000
Trip generation rate during the off-peak hour	λ	pax/hr	30,000
Loading time of the peak period	T	hr	1
Bus free-flow speed	u_b	km/hr	40
Average walking speed	w	km/hr	2
Time lost per stop	ω	sec	30
Boarding and alighting time per passenger	ω'	sec/pax	1
Unit dedicated-infrastructure cost	$\pi_{L,db}$	$/km-hr	90
Unit mixed-infrastructure cost	$\pi_{L,mb}$	$/km-hr	9
Unit vehicle cost	π_B	$/veh-hr	40
Unit distance cost	π_V	$/veh-km	2
Unit mileage cost	π_D	$/km	0.3
Value of time	α	$/pax-hr	20
Minimum headway	$H_{\min}$	min	3
Bus passenger capacity	C	pax/veh	150
Average car waiting time	W_c	min	7
Parameter of the Logit model	θ	$\emptyset$	1.5

4.2. Case study and analyzed scenarios

Here we investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network configuration and passenger mode choice (i.e. traffic performance). We do this by applying the previously described framework to the Barcelona network. Table 2 summarizes the input parameters for this case study. The values are similar to those in the literature (see e.g. Daganzo, 2010; Estrada et al., 2011; Badia et al., 2014; 2016).

4.3. General results and the effects of demand intensity

To quantify the effects of demand intensity, we analyze two scenarios: SC-I, indicating the base case, with the trip generation rates given in Table 2; and SC-II, representing an increase in the trip generation rate (both during the peak and off-peak period) by 33%. Similarly to Ouyang et al. (2014), we explore four distinct trip length patterns, reflecting different spatial demand distributions: (i) a uniform city (Fig. 5a), where origins and destinations are uniformly distributed across the entire network; (ii) a mono-centric city (Fig. 5c), where most of the origins and destinations are located within the city center, resulting in shorter trip lengths compared to (i) (see Fig. 5d); (iii) a commuter city (Fig. 5e), where most of the origins and destinations are located in two regions positioned in the opposite corners of the network, resulting in longer trip lengths compared to (i) (see Fig. 5f); and (iv) a twin city (Fig. 5g), where most of the origins and destinations are located in two nearly-adjacent regions near the city center. Notice that, due to the rectangular structure of the street network, the E-W traffic has a longer trip length than the N-S one in all considered patterns, as shown by the tails of the trip length distributions (Figs. 5b, 5 d, 5 f, and 5 h). Also note that the trip length distribution spreads the most for the uniform trip length pattern, given that the origins and destinations are distributed uniformly all across the whole network. As the origins and destinations become concentrated in a certain region of the network, the range of trip lengths gets reduced.

Table 3 shows the optimal design parameters, operator and user costs, as well as the corresponding mode choice, for each demand scenario, type of user behavior, and trip length pattern. The headways are given in minutes, whereas the costs are expressed in hours. The results suggest that for longer trip length patterns and higher demand levels more users tend to choose the bus system. The reason for this is twofold. First, the mileage cost component (besides the in-vehicle time traveled) in the car user cost function (defined as in Daganzo, 2010; Estrada et al., 2011; Badia et al., 2016) linearly increases with an increase in the trip length (see Eq. 22). This, however, is not the case for the bus mode, as all the cost

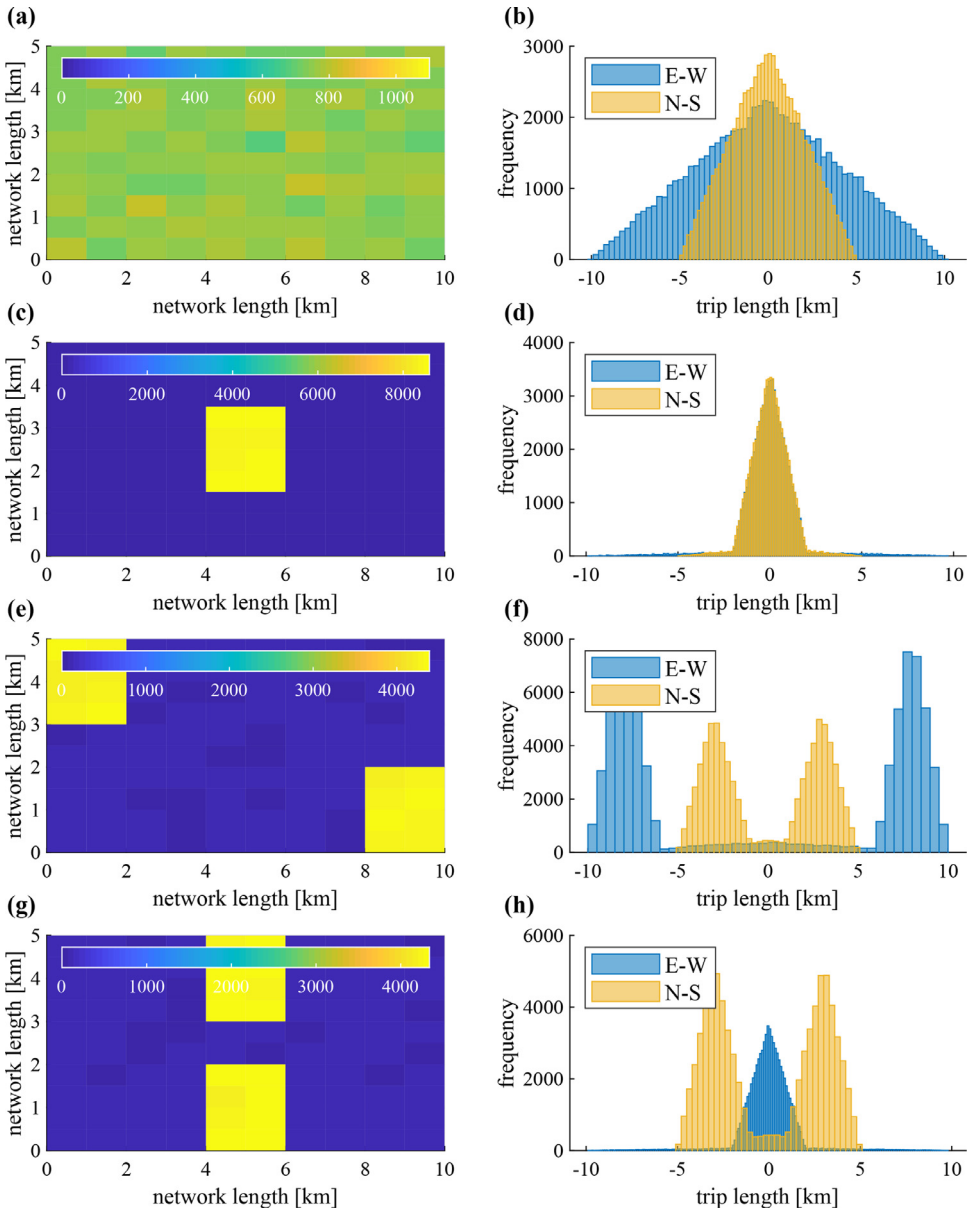

Fig. 5. Concentration of user origins and destinations, and the corresponding trip length pattern in case of: (a-b) uniform city; (c-d) a mono-centric city; (e-f) a commuter city; and (g-h) a twin city.

components (except for the in-vehicle time traveled) remain the same, i.e. they are independent of the user trip length. Second, by accounting for the interactions between the modes and their effects on the traffic performance, we observe that the system can benefit if more users select the bus mode, as otherwise the car delay would increase due to a higher car accumulation in the network. This, in turn, would also affect the bus system along mixed lanes, ultimately leading to overall lower network performance. To cope with increasing number of users in the system in scenario SC-II and ensure sufficient passenger capacity, the optimal bus network configuration requires a more frequent service compared to SC-I, as shown in Table 3. This holds true for both types of user behavior. Nevertheless, it is worth mentioning that the same pattern might not be observed in real-life scenarios in case the utility functions are significantly different from those used in the numerical experiments.

Notice from Table 3 that (i) both the user behavior and the trip length pattern do affect the optimal bus network configuration and passenger mode choice, and that (ii) the optimal design parameters are not constant across cardinal directions. This is an important insight, given that most scientific literature on the bus network design problem assumes, up to now, unique design parameters across the network and one particular type of user behavior and trip length pattern. Therefore,

Table 3
Optimal design parameters, operator cost, user cost, and the corresponding mode choice.

Scenario	User behavior	Trip pattern	$z_{s,x}$	$z_{s,y}$	$z_{l,x}$	$z_{l,y}$	H_x	H_y	$\zeta_{x,b,db}$	$\zeta_{y,b,db}$	Z_O	Z_U	Z	Pr(b)
SC-I	Type 1	(i)	2	3	2	2	6	5	0	0.5	0.065	0.538	0.603	0.36
		(ii)	1	2	3	3	8	6	0	0	0.061	0.357	0.418	0.36
		(iii)	3	3	1	2	3	3	1	1	0.112	0.826	0.938	0.46
		(iv)	1	3	3	2	7	4	0	0.9	0.081	0.479	0.560	0.40
	Type 2	(i)	2	4	4	1	5	3	0	1	0.068	0.491	0.559	0.42
		(ii)	1	4	7	1	6	4	0	0.7	0.061	0.311	0.372	0.42
		(iii)	3	3	1	2	3	4	1	1	0.099	0.792	0.891	0.48
		(iv)	1	4	7	1	6	3	0	1	0.064	0.437	0.501	0.44
SC-II	Type 1	(i)	2	3	2	2	5	3	0	0.9	0.065	0.525	0.590	0.38
		(ii)	1	2	3	3	5	5	0	0	0.058	0.347	0.405	0.38
		(iii)	3	4	1	1	3	3	1	1	0.099	0.838	0.937	0.48
		(iv)	1	4	3	1	6	3	0	1	0.078	0.462	0.540	0.42
	Type 2	(i)	2	3	4	1	5	3	0	1	0.062	0.484	0.546	0.42
		(ii)	1	3	8	1	5	4	0	1	0.061	0.295	0.356	0.44
		(iii)	3	4	1	1	3	3	1	1	0.095	0.803	0.898	0.50
		(iv)	1	4	6	1	5	3	0	1	0.057	0.432	0.489	0.44

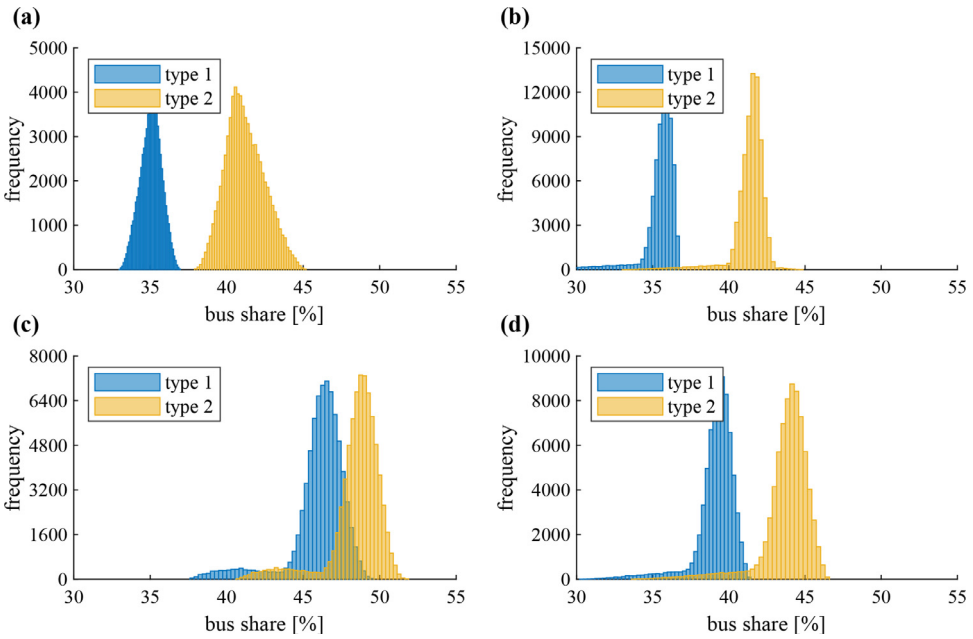

Fig. 6. Mode choice distribution for the optimal bus network design in scenario SC-I in case of: (a) uniform city; (b) a mono-centric city; (c) a commuter city; and (d) a twin city.

in the following, we demonstrate how important is to relax these assumptions. In particular, we quantify the value of each of the proposed methodological extensions including: (i) modeling of mode choice at the trip length level; (ii) considering heterogeneous types of user behavior; (iii) considering heterogeneous trip length patterns; (iv) accounting for the complex modal interactions; and (v) considering heterogeneous design parameters across cardinal directions.

4.4. Value of the proposed methodological extensions

4.4.1. Value of modeling the mode choice at the trip length level

In the previous subsection, we showed that the probability of choosing any given mode depends not only on the demand intensity, but also the type of user behavior. A closer look into the mode choice distribution reveals that this probability is not constant across the user trip lengths, but follows certain distribution (Fig. 6). This distribution is not unique, but varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Interestingly, the distribution for Type 2 behavior shows higher deviation and is shifted towards the right (i.e. higher number of users choose to use the bus system) compared to Type 1. The reason for this is that the bus system designed according to Type 2 behavior provides higher level of service, as elaborated further below. Note that, since the resulting mode choice distributions for both demand scenarios exhibit a similar shape, for brevity, we only show that of SC-I.

Table 4
Effects of user behavior on the cost functions.

Scenario	Trip pattern	ΔZ_O	ΔZ_U	ΔZ
SC-I	(i)	+4.46%	-8.70%	-7.28%
	(ii)	-1.02%	-12.74%	-11.02%
	(iii)	-11.62%	-4.27%	-5.15%
	(iv)	-21.24%	-8.90%	-10.69%
SC-II	(i)	-5.15%	-7.62%	-7.34%
	(ii)	+6.89%	-14.81%	-11.72%
	(iii)	-3.30%	-4.22%	-4.12%
	(iv)	-26.85%	-6.49%	-9.43%

Table 5
Effects of trip length patterns on the cost functions.

Scenario	User behavior	Trip pattern	ΔZ_O	ΔZ_U	ΔZ
SC-I	Type 1	(ii)	-5.48%	-0.55%	-1.31%
		(iii)	+109.93%	-9.89%	-3.30%
		(iv)	+38.46%	-4.67%	-0.15%
	Type 2	(ii)	-10.44%	-1.22%	-2.86%
		(iii)	+57.35%	-8.04%	-3.58%
		(iv)	-1.36%	-1.97%	-1.89%
SC-II	Type 1	(ii)	-11.68%	+0.39%	-1.52%
		(iii)	+79.41%	-10.59%	-5.64%
		(iv)	+31.48%	-4.33%	-0.42%
	Type 2	(ii)	+3.88%	-3.66%	-2.44%
		(iii)	+66.09%	-10.50%	-5.93%
		(iv)	-3.87%	-1.60%	-1.87%

4.4.2. Value of considering heterogeneous types of user behavior

For quantifying the importance of considering heterogeneous types of user behavior, we compare the optimal values of the cost functions for Type 1 and Type 2 behavior, such that $\Delta Z_O = (Z_{2,O} - Z_{1,O})/Z_{1,O}$, $\Delta Z_U = (Z_{2,U} - Z_{1,U})/Z_{1,U}$, and $\Delta Z = (Z_2 - Z_1)/Z_1$. The results of this comparison are given in Table 4. We observe that for all trip length patterns users can benefit (according to the cost functions used here) if they are willing to adjust the number of transfers to minimize the walking distance at the origin and the destination. This is consistent with empirical results from the city of Barcelona (Badia et al., 2017). Potential improvements can range from 4.2% to 14.8%, depending on the demand level and prevailing trip length pattern (see Table 4).

To motivate the users to use the transfers, i.e. provide a well-designed transfer-based network, the user waiting time needs to be minimized. This is indicated by the optimal decision variables in Table 3, where we can see that the optimal bus network configuration for Type 2 behavior provides higher bus frequency (i.e. lower headways in the two directions for almost all types of cities in both tested scenarios) compared to Type 1. Consequently, the operator cost might increase for some trip length patterns (e.g. uniform city in SC-I or mono-centric city in SC-II), as shown in Table 4. However, since the share of the operator cost within the total system cost is substantially lower than that of the user cost, the total cost function still gets reduced if the network is designed according to Type 2 user behavior. The improvements can range from 5.1% to 11.7% (see Table 4).

4.4.3. Value of considering heterogeneous trip length patterns

Here we quantify the importance of considering heterogeneous trip length patterns by comparing the costs obtained when the bus network parameters are optimally determined for a given trip length pattern, with those obtained if we were to assume uniformly distributed origins and destinations (Table 5). A negative value means that accounting for the actual trip length distribution does lead to lower costs. As expected, such lower total costs are achieved for both types of user behavior, with a similar magnitude of improvements. The improvements tend to increase with an increase in the demand level, especially for a commuter city (i.e. longest trip length pattern) that exhibits the highest improvement (5.9% for Type 2; 5.6% for Type 2) in the total cost function. In contrast, the lowest improvements (0.1% for Type 1; 1.9% for Type 2) are acquired for a twin city.

It is worth noting that in some cases the operator cost increases when we use the actual trip length distribution, especially for a commuter city. The reason for this is twofold. First, the optimal bus network design for a uniform city does not satisfy the constraint for the maximum bus occupancy (Eq. 11) when applied to a commuter city, thus cannot be considered to be feasible. From that perspective, an increase in the operator cost (imposed by considering the actual trip length distribution) for a commuter city is justified by ensuring that the bus capacity is sufficient to serve the maximum bus occupancy during the peak hour. Second, a commuter city requires a bus network with a higher spatial coverage and more frequent bus service compared to the uniform one. Such a requirement, however, ensures that the level of service provided by the bus system is competitive with that of the automobile at a reasonable cost. This is confirmed with the user cost function, which

Table 6
Effects of modal interactions on the cost functions.

Scenario	User behavior	Trip pattern	ΔZ_O	ΔZ_U	ΔZ
SC-I	Type 1	(i)	+123.13%	-7.15%	-0.92%
		(ii)	+110.91%	-11.17%	-2.90%
		(iii)	+86.47%	-10.87%	-4.95%
		(iv)	+106.08%	-11.71%	-3.71%
	Type 2	(i)	+111.59%	-9.76%	-3.01%
		(ii)	+89.49%	-13.98%	-5.54%
		(iii)	+68.76%	-12.04%	-7.09%
		(iv)	+61.59%	-16.46%	-10.95%
SC-II	Type 1	(i)	+96.99%	-7.49%	-1.72%
		(ii)	+74.14%	-10.64%	-3.98%
		(iii)	+46.70%	-11.35%	-7.52%
		(iv)	+94.44%	-12.30%	-4.76%
	Type 2	(i)	+101.26%	-11.47%	-5.47%
		(ii)	+102.59%	-18.07%	-8.7%
		(iii)	+46.66%	-12.87%	-8.98%
		(iv)	+46.11%	-14.24%	-9.91%

indeed gets reduced when we design the bus network for the actual trip length pattern. Consequently, the total system cost also gets reduced, given that the share of the user costs within the total system cost is substantially higher than that of the operator cost, as we stated before. In other words, increases in the operator cost are compensated by a much more efficient level of service from the user perspective, leading to overall lower costs.

4.4.4. Value of accounting for the complex modal interactions

So far, we have discussed how the demand intensity, user behavior, and trip length patterns affect the optimal bus network configuration and quantified the value of considering these factors. To realistically model both, the car and the bus system, we employed the 3D-MFD to take into account the interactions between the modes and their impact on the traffic performance. These interactions, however, are ignored in most scientific literature on the bus network design problem, which typically assumes that the travel times (or equivalently, the bus speeds) are independent of the design parameters. It is therefore the purpose of this subsection to quantify the value of considering the complex modal interactions in the proposed modeling framework.

We do this by comparing the results of the proposed approach to those obtained when the optimal design parameters for the simplified problem (i.e. considering the bus system only and using the average car speed during the peak period as the bus cruising speed along mixed lanes, as in Daganzo (2010) and Estrada et al. (2011)) are incorporated into the proposed framework. To make a fair comparison between the two approaches, we use the same demand level for the bus mode (determined from the mode choice in the original problem). The results of this comparison are shown in Table 6. A negative value means that a proper modeling of the complex bi-modal interactions does lead to lower costs. As expected, such lower total costs are achieved for both types of user behavior and all trip length patterns. Notice that the improvements in the total system cost increase with an increase in the demand level, especially for Type 1 behavior. These improvements come from the significant improvements (up to 18%) in the user cost function. That being said, we also observe an increase in the operator cost. This is because the optimal solution of the simplified problem produces a bus network design that operates in a fully mixed (with car traffic) environment across all cardinal directions, with a lower service frequency compared to that of the original problem. Consequently, once the solution of the simplified approach is incorporated into the proposed framework, the user waiting time increase and the occupancy rates become close or equal to the bus capacity (indicating lower level of service) for all demand levels, types of user behavior, and trip length patterns. In contrast, by accounting for the effects of the bus system on private car users, and vice versa, especially for higher demand levels, the proposed framework leads to a bus network design with higher percentage of dedicated bus lines. Due to reduced interactions between the modes, such design allows to increase the bus frequency along dedicated bus segments, improving thereby the level of service (reducing the passenger waiting time and bus occupancy) compared to the simplified approach. This indicates the value of the proposed modeling framework in capturing the necessary factors for determining the optimal bus network configuration, while taking into account multiple transport modes.

4.4.5. Value of considering heterogeneous design parameters

The results in Table 3 indicate that the optimal bus network configuration for different trip length patterns does not always provide a fully dedicated or mixed (with car traffic) bus system across the entire network (i.e. all cardinal directions), as it was the case in previous studies (see e.g. Daganzo, 2010; Estrada et al., 2011; Badia et al., 2016; Amirgholy et al., 2017). In fact, most of the decision variables vary along E-W and N-S. This holds true for both types of user behavior. Hence, it is worth quantifying the importance of considering a heterogeneous bus network design across cardinal directions.

We perform this analysis by comparing the results of the proposed approach to those obtained when the optimal design parameters for the simplified problem (i.e. considering the same stop spacing, bus headway, and lane allocation across all

Table 7
Effects of heterogeneous design parameters across cardinal directions on the cost functions.

Scenario	User behavior	Trip pattern	ΔZ_O	ΔZ_U	ΔZ
SC-I	Type 1	(i)	-12.60%	-0.01%	-0.63%
		(ii)	-12.60%	+0.04%	-0.87%
		(iii)	-10.38%	+0.06%	-0.54%
		(iv)	-8.05%	+0.09%	-0.45%
	Type 2	(i)	-11.05%	+0.04%	-0.45%
		(ii)	-11.05%	-0.00%	-0.68%
		(iii)	-9.34%	+0.05%	-0.42%
		(iv)	-5.52%	+0.04%	-0.26%
SC-II	Type 1	(i)	-6.94%	-0.07%	-0.41%
		(ii)	-6.94%	+0.09%	-0.42%
		(iii)	-9.02%	-0.01%	-0.53%
		(iv)	-7.13%	-0.08%	-0.53%
	Type 2	(i)	-8.86%	+0.15%	-0.21%
		(ii)	-5.96%	-0.05%	-0.36%
		(iii)	-8.94%	+0.01%	-0.44%
		(iv)	-6.25%	+0.07%	-0.25%

cardinal directions, as in Estrada et al., 2011) are incorporated into the proposed framework. The results of this comparison are shown in Table 7. Similarly to the previous analysis, a negative value means that a heterogeneous bus network design does lead to lower costs. We observe that the operator costs can indeed be reduced in both tested scenarios, for both types of user behavior and all trip length patterns, if we allow the decision variables to vary across cardinal directions. Potential improvements can range from 5.5% to 12.6%. Notice that there are almost no effects on the user cost function, resulting in marginal improvements in the total system cost (see Table 7). This shows that the proposed framework provides more flexibility for the bus system to serve the passenger demand while reducing the operator cost compared to the existing approaches.

5. Conclusions

In this study we investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network design, as well as the effects that the bus network structure might have on the traffic performance and passenger mode choice. For this purpose, we introduce several major extensions to the classical problem formulation. First, we consider two types of user behavior when selecting bus routes: (i) minimization of the number of transfers (Type 1); and (ii) minimization of the walking distance (Type 2). Second, we propose a new concept for considering trip length heterogeneity related to different origin-destination trips in the network. Third, we explicitly account for mixed traffic (i.e. mutual influence of cars and buses in the network) and mode choice at the trip length level. Forth, we use a dynamic description of the peak hour demand by distinguishing between the loading and unloading phase. This way, we are able to achieve a more realistic modeling of network congestion, hence determine a mean speed that depends on the bus network configuration, demand profile, and passenger mode choice.

Overall, the contributions of this research are sixfold: (i) we account for spatially non-uniform network topology and distribution of the passenger demand across cardinal directions; (ii) we use the distribution of the user trip lengths as an intermediate level of abstraction to determine the optimal design parameters. Such a level of abstraction allows not only to account for different trip length patterns (per and across cardinal directions) for the same level of passenger demand, but also to capture more accurately the network loading for all modes; (iii) we consider the complex modal interactions and passenger mode choice, thus include the travel costs for both, the bus and the car mode, into the objective function. This way, we optimize the performance of the whole network while taking into account multiple transport modes; (iv) we consider the influence of the passenger demand and network topology on the traffic performance, which, in turn, affects the passenger mode choice (i.e. demand). In other words, we use a full feedback loop to model all aforementioned dependencies; (v) we consider heterogeneous design parameters across cardinal directions, incorporating different lane allocations into the bus network design problem and accounting for their different influences on car traffic; and (vi) we investigate the effects of demand intensity, user behavior, and trip length patterns on the optimal bus network design; all in an analytical manner.

Numerical experiments performed for different trip length patterns, reflecting different spatial demand distributions, show that all the tested factors do affect the optimal bus network configuration and passenger mode choice. Results reveal that the probability of choosing any given mode is not constant across the user trip lengths, i.e. it follows certain distribution. This distribution is not unique, but varies across the trip length patterns, indicating the importance of modeling the mode choice at the trip length level. Moreover, the results indicate that a well-designed transfer-based bus network, i.e. a bus network designed according to Type 2 behavior, leads to lower costs and attracts more users for the same level of passenger demand compared to that of Type 1. This finding is consistent with empirical results from the city of Barcelona (Badia et al., 2017). Finally, we demonstrate the significance of addressing simplifications made in previous studies. From

that perspective, we show that the bus network parameters optimally determined for a given trip length pattern lead to lower costs compared to those obtained if we were to assume uniformly distributed origins and destinations. More importantly, the optimal bus network design determined for the uniform trip pattern does not satisfy the constraint for the maximum bus occupancy when applied to other trip length patterns, thus cannot be considered to be feasible. In other words, accounting for the actual trip length distribution allows the operator to design the bus network in accordance to the prevailing demand pattern. A comparison with a simplified approach that considers the bus system only and uses the average car speed during the peak period as the bus cruising speed along mixed lanes (as in Daganzo, 2010; Estrada et al., 2011) reveals the value of accounting for the complex modal interactions, especially for higher demand levels. Last, we show that by considering heterogeneous design parameters across cardinal directions we provide more flexibility for the bus system to serve the passenger demand while reducing the operator cost compared to the existing approaches.

It is also worth mentioning that, although the real networks are more complicated, the formulated design problem for simplified networks gives a fast global assessment of the key design variables (range of bus headways, overall fraction of dedicated bus lanes, range of stop and line spacings per cardinal direction) that should be further analyzed at the operational level when applied to real networks. That being said, we acknowledge that the proposed model might not be as parsimonious as those in the previous studies. Nevertheless, it still maintains the analytical structure of the problem while addressing important assumptions in the current state of the art, shown to have a significant impact, as we stated before.

Declaration of Competing Interest

No.

Acknowledgments

This work was supported by the Swiss National Science Foundation (SNSF) under the project name DIPLOMAT, contract 205121_165644. Igor Dakic acknowledges funding support by the Swiss National Science Foundation (SNSF) Doc.Mobility with project number P1EZP2_188003. Ludovic Leclercq acknowledges funding support by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 646592 MAGnUM project). Monica Menendez acknowledges support by the NYUAD Center for Interacting Urban Networks (CITIES), funded by Tamkeen under the NYUAD Research Institute Award CG001 and by the Swiss Re Institute under the Quantum Cities TM initiative.

Appendix A. Proofs

Result 1. The expected walking time at the origin and the destination for Type 2 user behavior is given by Eq. 27:

$$A_{2,b}(\boldsymbol{D}) = \begin{cases} (6l_x l_y + 6l_x s_x - 2l_y^2 - 3l_y s_x + 3l_y s_y)/(12 l_x w), & \text{if } l_x \geq l_y, \\ (6l_x l_y + 6l_y s_y - 2l_x^2 - 3l_x s_y + 3l_x s_x)/(12 l_y w), & \text{if } l_x < l_y. \end{cases}$$

Proof. Recall that for this type of behavior we assume users to choose the closest origin and destination stops such that they minimize the walking distance. Consequently, how the users access/exit the bus system (using a vertical or a horizontal bus line) depends on where their origin/destination is located within a rectangle defined by the horizontal and vertical line spacings (see Fig. 7). This, in turn, defines the walking distance (thus the access time), as well as the number of transfers (thus the waiting time). From that perspective, we can distinguish four regions within the rectangle, as shown in Fig. 7. Note that if the origin/destination falls within region (1) or (3), the users access/exit the bus system at the stop closest to their origin/destination along the nearest horizontal bus line. Otherwise, they access/exit the bus system at the closest stop along the nearest vertical bus line. As the shape of the rectangle and the corresponding regions depends on the ratio between l_x and l_y, we consider the following two cases:

Case 1. ($l_x \geq l_y$): regions (1) and (3) consist of two subregions: a rectangle with sides $l_x - l_y$ and $l_y/2$; and two right triangles with legs $l_y/2$ and $l_y/2$. The first subregion has the area of $P_{\square}^{(1)} = (l_x - l_y)l_y/2$ and the associated average walking distance of $(1/2)(l_y/2) = l_y/4$ along N-S and $(1/2)(s_x/2) = s_x/4$ along E-W directions, i.e. the average walking distance is $A_{\square}^{(1)} = (l_y/4 + s_x/4)/w$. The area of the second subregion (combining the two triangles) is $P_{\triangle}^{(1)} = l_y^2/4$, with the associated average walking distance of $(1/3)(l_y/2) = l_y/6$ along N-S and $(1/2)(s_x/2) = s_x/4$ along E-W directions, i.e. $A_{\triangle}^{(1)} = (l_y/6 + s_x/4)/w$. Similarly, regions (2) and (4) have each the area of $P_{\triangle}^{(2)} = l_y^2/4$, with the associated average walking distance of $(1/3)(l_y/2) = l_y/6$ along E-W and $(1/2)(s_y/2) = s_y/4$ along N-S directions, i.e. $A_{\triangle}^{(2)} = (l_y/6 + s_y/4)/w$. Then, assuming that the expected access and egress distances traveled are the same (similarly to Daganzo, 2010; Estrada et al., 2011), the expected walking time can be computed by considering all possible

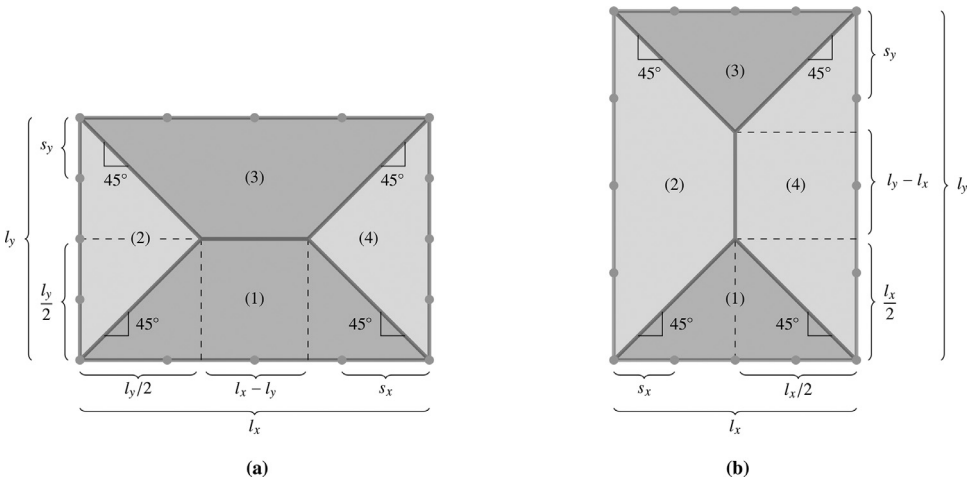

Fig. 7. Regions defining potential locations of an origin/destination within a rectangle defined by sides l_x and l_y in case: (a) $l_x \geq l_y$; and (b) $l_x < l_y$.

combinations of the subregions in which the origin/destination might be located:

$$A_{2,b}(\boldsymbol{D}) = 2\left(A_{\square}^{(1)} \cdot \frac{P_{\square}^{(1)}}{P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}} + A_{\Delta}^{(1)} \cdot \frac{P_{\Delta}^{(1)}}{P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}} + A_{\Delta}^{(2)} \cdot \frac{P_{\Delta}^{(2)}}{P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}}\right)$$
$$= \left(6l_x l_y + 6l_x s_x - 2l_y^2 - 3l_y s_x + 3l_y s_y\right)/(12 l_x w).$$

Case 2. ($l_x < l_y$): regions (1) and (3) have the area of $P_{\Delta}^{(1)} = l_x^2/4$ and the associated average walking distance of $(1/3)(l_x/2) = l_x/6$ along N-S and $(1/2)(s_x/2) = s_x/4$ along E-W directions, i.e. $A_{\Delta}^{(1)} = (l_x/6 + s_x/4)/w$. On the other hand, regions (1) and (3) consist of two subregions (similarly to regions (1) and (3) in Case 1): a rectangle with sides $l_y - l_x$ and $l_x/2$; and two right triangles with legs $l_x/2$ and $l_x/2$. The first subregion has the area of $P_{\square}^{(2)} = (l_y - l_x)l_x/2$ and the associated average walking distance of $(1/2)(l_x/2) = l_y/4$ along E-W and $(1/2)(s_y/2) = s_y/4$ along N-S directions, i.e. $A_{\square}^{(2)} = (l_x/4 + s_y/4)/w$. The area of the second subregion (combining the two triangles) is $P_{\Delta}^{(2)} = l_x^2/4$, with the associated average walking distance of $(1/3)(l_x/2) = l_x/6$ along E-W and $(1/2)(s_y/2) = s_y/4$ along N-S directions, i.e. $A_{\Delta}^{(2)} = (l_x/6 + s_y/4)/w$. Similarly to the previous case, the expected walking time is given as:

$$A_{2,b}(\boldsymbol{D}) = 2\left(A_{\Delta}^{(1)} \cdot \frac{P_{\Delta}^{(1)}}{P_{\Delta}^{(1)} + P_{\square}^{(2)} + P_{\Delta}^{(2)}} + A_{\square}^{(2)} \cdot \frac{P_{\square}^{(2)}}{P_{\Delta}^{(1)} + P_{\square}^{(2)} + P_{\Delta}^{(2)}} + A_{\Delta}^{(2)} \cdot \frac{P_{\Delta}^{(2)}}{P_{\Delta}^{(1)} + P_{\square}^{(2)} + P_{\Delta}^{(2)}}\right)$$
$$= \left(6l_x l_y + 6l_y s_y - 2l_x^2 - 3l_x s_y + 3l_x s_x\right)/(12 l_y w).$$

□

Result 2. The expected number of transfers per trip for Type 2 user behavior is given by Eq. 28:

$$\varphi_2(\boldsymbol{D}) = \begin{cases} \Pr(\varphi > 0)(4l_x^2 l_y^2 - 2l_x l_y^3 + l_y^4)/(2l_x^2 l_y^2), & \text{if } l_x \geq l_y, \\ \Pr(\varphi > 0)(4l_x^2 l_y^2 - 2l_x^3 l_y + l_x^4)/(2l_x^2 l_y^2), & \text{if } l_x < l_y. \end{cases}$$

Proof. As mentioned before, the expected number of transfers depends on the location of the user origin and the destination within a rectangle defined by sides l_x and l_y. From that perspective, if the origin is in region (1) or (3) and the destination is in region (1) or (3), but not along the same corridor as the origin, users need to make two transfers (from a horizontal bus line along which they access the system to a vertical one; and from a vertical to another horizontal bus line from which they reach the destination). Likewise, if the origin is in region (2) or (4) and the destination is in region (2) or (4), but not along the same corridor as the origin, users also need to make two transfers (from a vertical bus line along which they access the system to a horizontal one; and from a horizontal to another vertical bus line from which they reach the destination). If the origin is in region (1) or (3) and the destination is in region (2) or (4), or vice versa, users make one transfer. If both, the origin and the destination, are along the same corridor (a horizontal corridor for regions regions (1) and (3), and a vertical corridor for regions (2) and (4)), users do not make any transfer. This leads to the following formulations for the expected number of transfers for each rectangular shape shown in Fig. 7. Recall that the probability that both, the origin and the destination, are along the same corridor is equivalent to the probability of making zero transfer, $\Pr(\varphi = 0) = (l_y\phi_x + l_x\phi_y - l_x l_y)/(\phi_x\phi_y)$.

Case 1. ($l_x \geq l_y$):

$$\begin{aligned}\varphi_2(\boldsymbol{D}) &= 0 \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)(1 \cdot \Pr(\varphi = 1) + 2 \cdot \Pr(\varphi = 2)) \\ &= \Pr(\varphi > 0)\left(1 \cdot \frac{2\left(P_{\square}^{(1)} + P_{\Delta}^{(1)}\right)P_{\Delta}^{(2)}}{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}\right)^2} + 2 \cdot \frac{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)}\right)^2 + \left(P_{\Delta}^{(2)}\right)^2}{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}\right)^2}\right) \\ &= \Pr(\varphi > 0)\left(4l_x^2 l_y^2 - 2l_x l_y^3 + l_y^4\right)/\left(2l_x^2 l_y^2\right)\end{aligned}$$

Case 2. ($l_x < l_y$):

$$\begin{aligned}\varphi_2(\boldsymbol{D}) &= 0 \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)(1 \cdot \Pr(\varphi = 1) + 2 \cdot \Pr(\varphi = 2)) \\ &= \Pr(\varphi > 0)\left(1 \cdot \frac{2P_{\Delta}^{(1)}\left(P_{\square}^{(2)} + P_{\Delta}^{(2)}\right)}{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}\right)^2} + 2 \cdot \frac{\left(P_{\Delta}^{(1)}\right)^2 + \left(P_{\square}^{(2)} + P_{\Delta}^{(2)}\right)^2}{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}\right)^2}\right) \\ &= \Pr(\varphi > 0)\left(4l_x^2 l_y^2 - 2l_x^3 l_y + l_x^4\right)/\left(2l_x^2 l_y^2\right)\end{aligned}$$

□

Result 3. The expected waiting time per user including that at the origin and all transfer stops for Type 2 user behavior is given by Eq. 29:

$$W_{2,b}(\boldsymbol{D}) = \begin{cases} \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)(H_x(8l_x^2 l_y^2 - 4l_x l_y^3 + l_y^4) + H_y(4l_x^2 l_y^2 + l_y^4))/(8l_x^2 l_y^2), & \text{if } l_x \geq l_y, \\ \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)(H_x(4l_x^2 l_y^2 + l_x^4) + H_y(8l_x^2 l_y^2 - 4l_x^3 l_y + l_x^4))/(8l_x^2 l_y^2), & \text{if } l_x < l_y. \end{cases}$$

Proof. Following the rationale for the expected number of transfers, we compute the expected waiting time as follows: (i) users with zero transfer wait on average half of the weighted bus headway in the network, i.e. $H(i) = \bar{H}/2 = (N_{b,x}H_x + N_{b,y}H_y)/(2(N_{b,x} + N_{b,y}))$; (ii) users with two transfers, the origin in region (1) or (3), and the destination in region (1) or (3) use two horizontal and one vertical bus line, thus wait on average $H(ii) = H_x + H_y/2$; (iii) users with two transfers, the origin in region (2) or (4), and the destination in region (2) or (4) use two vertical and one horizontal bus line, thus wait on average $H(iii) = H_y + H_x/2$; and (iv) users with one transfer, the origin in region (1) or (3), and the destination in region (2) or (4), or vice versa, use one horizontal and one vertical bus line, thus wait on average $H(iv) = H_x/2 + H_y/2$. As for the previous parameters of the user cost function for the bus mode, the expected waiting time is computed by multiplying these four groups of users by the corresponding probabilities (and in case of users with non-zero transfer by the probability of making a transfer $\Pr(\varphi > 0) = 1 - \Pr(\varphi = 0)$) for each rectangular shape shown in Fig. 7.

Case 1. ($l_x \geq l_y$):

$$\begin{aligned}W_{2,b}(\boldsymbol{D}) &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0))H(ii) \cdot \Pr(ii) + H(iii) \cdot \Pr(iii) + H(iv) \cdot \Pr(iv)) \\ &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)\left(H(ii) \cdot \frac{\left(P_{\square}^{(1)} + P_{?}^{(1)}\right)^2}{\left(P_{\square}^{(1)} + P_{?}^{(1)} + P_{?}^{(2)}\right)^2}\right. \\ &\quad \left. + H(iii) \cdot \frac{\left(P_{?}^{(2)}\right)^2}{\left(P_{\square}^{(1)} + P_{?}^{(1)} + P_{?}^{(2)}\right)^2} + H(iv) \cdot \frac{2\left(P_{\square}^{(1)} + P_{?}^{(1)}\right)P_{?}^{(2)}}{\left(P_{\square}^{(1)} + P_{?}^{(1)} + P_{?}^{(2)}\right)^2}\right) \\ &= \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)\left(H_x\left(8l_x^2 l_y^2 - 4l_x l_y^3 + l_y^4\right) + H_y\left(4l_x^2 l_y^2 + l_y^4\right)\right)/\left(8l_x^2 l_y^2\right)\end{aligned}$$

Case 2. ($l_x < l_y$):

$$\begin{aligned}W_{2,b}(\boldsymbol{D}) &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)(H(ii) \cdot \Pr(ii) + H(iii) \cdot \Pr(iii) + H(iv) \cdot \Pr(iv)) \\ &= H(i) \cdot \Pr(\varphi = 0) + \Pr(\varphi > 0)\left(H(ii) \cdot \frac{\left(P_{\Delta}^{(1)}\right)^2}{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}\right)^2}\right. \\ &\quad \left. + H(iii) \cdot \frac{\left(P_{\square}^{(2)} + P_{\Delta}^{(2)}\right)^2}{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}\right)^2} + H(iv) \cdot \frac{2P_{\Delta}^{(1)}\left(P_{\square}^{(2)} + P_{\Delta}^{(2)}\right)}{\left(P_{\square}^{(1)} + P_{\Delta}^{(1)} + P_{\Delta}^{(2)}\right)^2}\right) \\ &= \Pr(\varphi = 0)\bar{H}/2 + \Pr(\varphi > 0)\left(H_x\left(4l_x^2 l_y^2 + l_x^4\right) + H_y\left(8l_x^2 l_y^2 - 4l_x^3 l_y + l_x^4\right)\right)/\left(8l_x^2 l_y^2\right)\end{aligned}$$

□

Result 4. When the trip origins and destinations are uniformly and independently distributed across the network, Eq. 31 becomes equivalent to that of Daganzo (2010) for grid-like structures:

$$O_p = \begin{cases} (\Lambda/4)H_x/N_{x,b}, & \text{if } p \in \{eb, wb\}, \\ (\Lambda/4)H_y/N_{y,b}, & \text{if } p \in \{nb, sb\}. \end{cases}$$

Proof. It suffices to show that the maximum number of on-board passengers along E-W/N-S directions is across the vertical/horizontal cordon that bisects the city, i.e. $\beta = 1/2$ (see Fig. 2). For brevity, we only show this for the *eb* direction. Given the symetry of the network, derivations for other cardinal directions are done in an entirely analogous manner using the corresponding subscripts.

To be able to obtain the same result as Daganzo (2010), we need to use the same assumption that the abscissa and the ordinate of both, the trip origins (x_1, y_1) and the destinations (x_2, y_2), are distributed uniformly and independently on $[0, \phi_x]$ and $[-\phi_y, \phi_y]$, respectively. As abscissae x_1 and x_2 are independent uniforms, their difference $x_1 - x_2$ has a triangular distribution, hence the user trip length in the *eb* direction $\{\ell_x = x_1 - x_2 : \ell_x \geq 0\}$ has the following probability density function: $\{f(\ell_x) = (\phi_x - \ell_x)/\phi_x^2 : \ell_x \in [0, \phi_x]\}$. Furthermore, due to the independence, the joint probability density function becomes $f(\ell_x, \ell_y) = f(\ell_x) f(\ell_y)$. Finally, since Daganzo (2010) does not consider the mode choice, we also assume that $\Pr(b|\ell_x, \ell_y) = 1, \ \forall \ell_x, \ell_y$. It follows that the maximum number of on-board passengers in the *eb* direction is:

$$
\begin{aligned}
Q_{eb} &= \max_{\beta\in[0,1]} \left\{ \int_0^{\phi_x} \int_{-\phi_y}^{\phi_y} \Pr(J_\beta(\ell_x)) \Lambda \Pr(b|\ell_x, \ell_y) f(\ell_x, \ell_y) \mathrm{d}\ell_y \mathrm{d}\ell_x \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \int_0^{\phi_x} \Pr(J_\beta(\ell_x))(\phi_x - \ell_x) \mathrm{d}\ell_x \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \left(\int_0^{\min\{\beta\phi_x, (1-\beta)\phi_x\}} \frac{\ell_x}{\phi_x - \ell_x} (\phi_x - \ell_x) \mathrm{d}\ell_x \right.\right. \\
&\quad + \int_{\min\{\beta\phi_x, (1-\beta)\phi_x\}}^{\max\{\beta\phi_x, (1-\beta)\phi_x\}} \frac{\min\{\beta\phi_x, (1-\beta)\phi_x\}}{\phi_x - \ell_x} (\phi_x - \ell_x) \mathrm{d}\ell_x \\
&\quad \left.\left. + \int_{\max\{\beta\phi_x, (1-\beta)\phi_x\}}^{\phi_x} (\phi_x - \ell_x) \mathrm{d}\ell_x \right) \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \left(\frac{1}{2} (\max\{\beta\phi_x, (1-\beta)\phi_x\})^2 \right.\right. \\
&\quad + \min\{\beta\phi_x, (1-\beta)\phi_x\} (\max\{\beta\phi_x, (1-\beta)\phi_x\} - \min\{\beta\phi_x, (1-\beta)\phi_x\}) \\
&\quad \left.\left. + \phi_x (\phi_x - \max\{\beta\phi_x, (1-\beta)\phi_x\}) - \frac{1}{2} (\phi_x^2 - (\max\{\beta\phi_x, (1-\beta)\phi_x\})^2) \right) \right\} \\
&= \max_{\beta\in[0,1]} \left\{ \frac{\Lambda}{\phi_x^2} \left(\min\{\beta\phi_x, (1-\beta)\phi_x\} \max\{\beta\phi_x, (1-\beta)\phi_x\} \right.\right. \\
&\quad - \frac{1}{2} (\min\{\beta\phi_x, (1-\beta)\phi_x\})^2 + \frac{1}{2}\phi_x^2 - \phi_x \max\{\beta\phi_x, (1-\beta)\phi_x\} \\
&\quad \left.\left. + \frac{1}{2} (\max\{\beta\phi_x, (1-\beta)\phi_x\})^2 \right) \right\}
\end{aligned}
$$

To complete the proof, let us consider the following two cases:

Case 1. $\beta\phi_x \leq (1-\beta)\phi_x$:

$$
\begin{aligned}
Q_{eb}(\beta) &= \frac{\Lambda}{\phi_x^2} \left(\beta\phi_x (1-\beta)\phi_x - \frac{1}{2}\beta^2\phi_x^2 + \frac{1}{2}\phi_x^2 - \phi_x(1-\beta)\phi_x + \frac{1}{2}(1-\beta)^2\phi_x^2 \right) \\
&= \Lambda(\beta - \beta^2)
\end{aligned}
$$

Case 2. $\beta\phi_x > (1-\beta)\phi_x$:

$$
\begin{aligned}
Q_{eb}(\beta) &= \frac{\Lambda}{\phi_x^2} \left((1-\beta)\phi_x \beta\phi_x - \frac{1}{2}(1-\beta)^2\phi_x^2 + \frac{1}{2}\phi_x^2 - \phi_x\beta\phi_x + \frac{1}{2}\beta^2\phi_x^2 \right) \\
&= \Lambda(\beta - \beta^2)
\end{aligned}
$$

Notice that both cases yield the same result. Therefore, we can compute the value of β for which the maximum value of $Q_{eb}(\beta)$ is reached by solving the following equation:

$$
\frac{\mathrm{d}Q_{eb}(\beta)}{\mathrm{d}\beta} = 0 \Leftrightarrow \Lambda(1 - 2\beta) = 0 \Leftrightarrow \beta = \frac{1}{2}
$$

It then follows that the maximum number of on-board passengers in the *eb* direction is $Q_{eb} = Q_{eb}(\beta = 1/2) = \Lambda/4$, which is the same result as that of Daganzo (2010). □

References

Ambühl, L., Loder, A., Zheng, N., Axhausen, K.W., Menendez, M., 2019. Approximative network partitioning for mfds from stationary sensor data. Transp. Res. Rec. 2673 (6), 94–103.

Amirgholy, M., Shahabi, M., Gao, H.O., 2017. Optimal design of sustainable transit systems in congested urban networks: a macroscopic approach. Transportation Research Part E: Logistics and Transportation Review 103, 261–285.

Ampountolas, K., Zheng, N., Geroliminis, N., 2017. Macroscopic modelling and robust control of bi-modal multi-region urban road networks. Transportation Research Part B: Methodological 104, 616–637.

Badia, H., Argote-Cabanero, J., Daganzo, C.F., 2017. How network structure can boost and shape the demand for bus transit. Transportation Research Part A: Policy and Practice 103, 83–94.

Badia, H., Estrada, M., Robusté, F., 2014. Competitive transit network design in cities with radial street patterns. Transportation Research Part B: Methodological 59, 161–181.

Badia, H., Estrada, M., Robusté, F., 2016. Bus network structure and mobility pattern: amonocentric analytical approach on a grid street layout. Transportation Research Part B: Methodological 93, 37–56.

Batista, S.F.A., Cantelmo, G., Menendez, M., Antoniou, C., 2020. The calculation of the virtual set of trips for the calibration of aggregated traffic models. 3rd Symposium on Management of Future Motorway and Urban Traffic Systems.

Bhat, C.R., Guo, J., 2004. A mixed spatially correlated logit model: formulation and application to residential choice modeling. Transportation Research Part B: Methodological 38 (2), 147–168.

Byrne, B.F., 1975. Public transportation line positions and headways for minimum user and system cost in a radial case. Transp. Res. 9 (2–3), 97–102.

Chen, H., Gu, W., Cassidy, M.J., Daganzo, C.F., 2015. Optimal transit service atop ring-radial and grid street networks: a continuum approximation design method and comparisons. Transportation Research Part B: Methodological 81, 755–774.

Daganzo, C.F., 2010. Structure of competitive transit networks. Transportation Research Part B: Methodological 44 (4), 434–446.

Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011. Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability. Transportation Research Part B: Methodological 45 (1), 278–288.

Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transportation Research Part B: Methodological 42 (9), 771–781.

Dakic, I., Ambühl, L., Schümperlin, O., Menendez, M., 2020. On the modeling of passenger mobility for stochastic bi-modal urban corridors. Transportation Research Part C: Emerging Technologies 113, 146–163.

Dakic, I., Menendez, M., 2018. On the use of lagrangian observations from public transport and probe vehicles to estimate car space-mean speeds in bi-modal urban networks. Transportation Research Part C: Emerging Technologies 91, 317–334.

Dakic, I., Yang, K., Menendez, M., 2019. Evaluating the effects of passenger occupancy dynamics on a bi-modal perimeter control. TRB Annual Meeting Online. Transportation Research Board.

Dakic, I., Yang, K., Menendez, M., Chow, J. Y., 2021. On the design of an optimal flexible bus dispatching system with modular bus units: Using the three--dimensional macroscopic fundamental diagram. Transportation Research Part B: Methodological 148, 38–59.

Estrada, M., Roca-Riu, M., Badia, H., Robusté, F., Daganzo, C.F., 2011. Design and implementation of efficient transit networks: procedure, case study and validity test. Procedia - Social and Behavioral Sciences 17, 113–135.

Fan, W., Mei, Y., Gu, W., 2018. Optimal design of intersecting bimodal transit networks in a grid city. Transportation Research Part B: Methodological 111, 203–226.

Gayah, V.V., Daganzo, C.F., 2011. Clockwise hysteresis loops in the macroscopic fundamental diagram: an effect of network instability. Transportation Research Part B: Methodological 45 (4), 643–655.

Gayah, V.V., Gao, X.S., Nagle, A.S., 2014. On the impacts of locally adaptive signal control on urban network stability and the macroscopic fundamental diagram. Transportation Research Part B: Methodological 70, 255–268.

Ge, Q., Menendez, M., 2017. Extending morris method for qualitative global sensitivity analysis of models with dependent inputs. Reliability Engineering & System Safety 162, 28–39.

Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transportation Research Part C: Emerging Technologies 42, 168–181.

Girault, J.-T., Gayah, V.V., Guler, S.I., Menendez, M., 2016. An exploratory analysis of signal coordination impacts on the macroscopic fundamental diagram. Transp. Res. Rec. 2560, 36–46.

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L., 2008. Understanding individual human mobility patterns. Nature 453 (7196), 779–782.

He, H., Menendez, M., Guler, S.I., 2018. Analytical evaluation of flexible-sharing strategies on multimodal arterials. Transportation Research Part A: Policy and Practice 114, 364–379.

He, H., Yang, K., Liang, H., Menendez, M., Guler, S.I., 2019. Providing public transport priority in the perimeter of urban networks: a bimodal strategy. Transportation Research Part C: Emerging Technologies 107, 171–192.

Holroyd, E., 1967. The optimum bus service: a theoretical model for a large uniform urban area. In: Proceedings of the Third International Symposium on the Theory of Traffic FlowOperations Research Society of America.

Ibarra-Rojas, O., Delgado, F., Giesen, R., Muñoz, J., 2015. Planning, operation, and control of bus transport systems: a literature review. Transportation Research Part B: Methodological 77, 38–75.

Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic fundamental diagrams: a cross-comparison of estimation methods. Transportation Research Part B: Methodological 62, 1–12.

Levinson, D., 2012. Network structure and city size. PLoS ONE 7 (1), e29721.

Little, J.D., 1961. A proof for the queuing formula: l= λ w. Oper. Res. 9 (3), 383–387.

Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2017. Empirics of multi-modal traffic networks - Using the 3D macroscopic fundamental diagram. Transportation Research Part C: Emerging Technologies 82, 88–101.

Loder, A., Ambühl, L., Menendez, M., Axhausen, K.W., 2019. Understanding traffic capacity of urban networks. Sci. Rep. 9 (1), 1–10.

Loder, A., Dakic, I., Bressan, L., Ambühl, L., Bliemer, M.C., Menendez, M., Axhausen, K.W., 2019. Capturing network properties with a functional form for the multi-modal macroscopic fundamental diagram. Transportation Research Part B: Methodological 129, 1–19.

Mühlich, N., Gayah, V.V., Menendez, M., 2015. Use of microsimulation for examination of macroscopic fundamental diagram hysteresis patterns for hierarchical urban street networks. Transp. Res. Rec. 2491 (1), 117–126.

Newell, G., 1979. Some issues relating to the optimal design of bus routes. Transportation Science 13 (1), 20–35.

Nikias, V., Guler, S.I., Menendez, M., 2016. Effects of bus operations on the traffic capacity of urban networks: a simulation study. Transportation Research Board 95th Annual Meeting.

Nourbakhsh, S.M., Ouyang, Y., 2012. A structured flexible transit system for low demand areas. Transportation Research Part B: Methodological 46 (1), 204–216.

Ortigosa, J., Gayah, V.V., Menendez, M., 2019. Analysis of one-way and two-way street configurations on urban grid networks. Transportmetrica B: transport dynamics 7 (1), 61–81.

Ortigosa, J., Menendez, M., Gayah, V.V., 2015. Analysis of network exit functions for various urban grid network configurations. Transportation Research Record: Journal of the Transportation Research Board 2491 (15–2886), 12–21. doi:10.3141/2491-02.

Ortigosa, J., Zheng, N., Menendez, M., N. Geroliminis, 2017. Traffic performance and road space allocation in multimodal urban networks with an MFD representation. 96th Transportation Research Board Annual Meeting (TRB), Washington D.C..

Ouyang, Y., Nourbakhsh, S.M., Cassidy, M.J., 2014. Continuum approximation approach to bus network design under spatially heterogeneous demand. Transportation Research Part B: Methodological 68, 333–344.
Paipuri, M., Leclercq, L., 2020. Bi-modal macroscopic traffic dynamics in a single region. Transportation research part B: methodological 133, 257–290.
Paipuri, M., Xu, Y., Gonzlez, M.C., Leclercq, L., 2020. Estimating mfds, trip lengths and path flow distributions in a multi-region setting using mobile phone data. Transportation Research Part C: Emerging Technologies 118, 102709.
Roca-Riu, M., Menendez, M., Dakic, I., Buehler, S., Ortigosa, J., 2020. Urban space consumption of cars and buses: an analytical approach. Transportmetrica B: Transport Dynamics 8 (1), 237–263.
Saade, N., Doig, J., Cassidy, M.J., 2018. Scheduling lane conversions for bus use on city-wide scales and in time-varying congested traffic. Transportation Research Part C: Emerging Technologies 95, 248–260.
Saade, N., Doig, J., Cassidy, M.J., 2018. Scheduling lane conversions for bus use on city-wide scales and in time-varying congested traffic. Transportation Research Part C: Emerging Technologies 95, 248–260.
Small, K., 2004. Urban transportation. concise encyclopedia of economics, 2nd.
Wirasinghe, S.C., Hurdle, V.F., Newell, G.F., 1977. Optimal parameters for a coordinated rail and bus transit system. Transportation Science 11 (4), 359–374.
Wu, K., Guler, S.I., 2018. Optimizing transit signal priority implementation along an arterial. Transp. Res. Rec. 2672 (20), 215–227.
Yang, H., Ke, J., Ye, J., 2018. A universal distribution law of network detour ratios. Transportation Research Part C: Emerging Technologies 96, 22–37.
Yildirimoglu, M., Petit, A., Geroliminis, N., Ouyang, Y., 2016. Bus service design under demand diversion and dynamic roadway congestion based on aggregated network models. TRB Annual Meeting Online. Transportation Research Board.
Zhang, F., Liu, W., 2019. Responsive bus dispatching strategy in a multi-modal and multi-directional transportation system: a doubly dynamical approach. Transportation Research Part C: Emerging Technologies.
Zhang, F., Zheng, N., Yang, H., Geroliminis, N., 2018. A systematic analysis of multimodal transport systems with road space distribution and responsive bus service. Transportation Research Part C: Emerging Technologies 96, 208–230.
Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Transportation Research Part B: Methodological 57, 326–341.