The effect of variability of urban systems characteristics in the network capacity

Nikolas Geroliminis *, Burak Boyacı

Urban Transport Systems Laboratory, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ABSTRACT

Recent experimental analysis has shown that some types of urban networks exhibit a low scatter reproducible relationship between average network flow and density, known as the macroscopic fundamental diagram (MFD). It has also been shown that heterogeneity in the spatial distribution of density can significantly decrease the network flow for the same value of density. Analytical theories have been developed to explore the connection between network structure and an MFD for urban neighborhoods with cars controlled by traffic signals. However these theories have been applied only in cities with deterministic values of topological and control variables for the whole network and by ignoring the effect of turns. In our study we are aiming to generate an MFD for streets with variable link lengths and signal characteristics and understand the effect of variability for different cities and signal structures. Furthermore, this variability gives the opportunity to mimic the effect of turning movements. Route or network capacity can be significantly smaller than the capacity of a single link, because of the correlations developed through the different values of offsets. The above analysis would not be possible using standard traffic engineering techniques. This will be a key issue in planning the signal regimes such a way that maximizes the network capacity and/or the density range of the maximum capacity.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Recent studies (Geroliminis and Sun, 2011b; Mazloumian et al., 2010; Daganzo et al., 2011) has shown that networks with heterogeneous distribution of link density exhibit network flows smaller than those that approximately meet homogeneity conditions (low spatial variance of link density), especially for congested conditions. Also, note that the scalability of flows from a series of links to large traffic networks is not a direct transformation. For example, route or network capacity can be significantly smaller than the capacity of a single link as this is expressed by sG/C (s is the saturation flow, G and C are the durations of green phase and cycle). This is because of the correlations between successive arterial links, the creation of spillback queues and the effect of offsets (Daganzo and Geroliminis, 2008). In case of long links, these effects are negligible and the propagation of traffic is much simpler. Nevertheless, congestion often occurs in the city centers with dense topology of short links.

At the link scale, traffic flows can be unpredictable or chaotic when a network is critically congested because of different driving behavior patterns, the effect of route choice, the fast dynamics of link travel times and origin–destination tables and the computational complexity (too many particles/cars). These observations make the development of global traffic management strategies, to improve mobility for a large signalized traffic network with a microscopic analysis, intractable. An alternative is a hierarchical control structure, where a network can be partitioned in homogeneous regions (with small spatial

variance of link density distribution) and optimal control methodologies can identify the inter-transfers between regions of a city to maximize the system output, as expressed by the number of trip endings. These policies can change the spatial distribution of congestion in such a way that the network outflow increases. This is a challenging task that requires knowledge on how the network flow for a region of a city changes as a function of topology, control and level of congestion.

The physical tool to advance this research is the macroscopic fundamental diagram (MFD) of urban traffic, which provides for some network regions a unimodal, low-scatter relationship between network vehicle density (veh/km) and network space-mean flow (veh/h). The first theoretical proposition of such a physical model was developed by Godfrey (1969), while similar approaches were also initiated by Herman and Prigogine (1979) and Daganzo (2007). The physical model of MFD was observed with dynamic features through empirical data in congested urban networks in Yokohama (Geroliminis and Daganzo, 2008). Other empirical or simulated analysis for MFDs with low or high scatter can be found in Buisson and Ladier (2009), Daganzo et al. (2011), Geroliminis and Sun (2011a), Gayah and Daganzo (2011), Courbon and Leclercq (2011), Ji et al. (2010), Saberi and Mahmassani, 2012 and others. Nevertheless, it is not obvious whether the MFDs would be universal or network-specific. More real-world experiments are needed to identify the types of networks and demand conditions, for which invariant MFDs with low scatter exist.

To evaluate topological or control-related changes of the network (e.g. a re-timing of the traffic signals or a change in infrastructure), Daganzo and Geroliminis (2008) and, Helbing (2009) have derived analytical theories for the urban fundamental diagram, using a density-based and a utilization-based approaches respectively. The first reference proved, using variational theory (Daganzo, 2005), that an MFD must arise for single-route networks with a fixed number of vehicles in circulation (periodic boundary conditions and no turns). The same reference also gives explicit formulae for the single-route MFD with deterministic topology, control and traffic characteristics (i.e. all intersections have common control patterns, the length of its links and their individual fundamental diagrams are all the same. The reference conjectured that these MFD formulae should approximately expected to hold for homogeneous, redundant networks with slow-changing demand. The methodology estimates the average speed and the maximum passing rate (rate that cars can overpass him) for a large number of observers moving forward or backward and stopping only at traffic lights. Then by considering that each observer can create a "cut" in the MFD, its shape is estimated as the lower envelope of all these cuts.

In this paper we provide several extensions and refinements of the analytical model for an MFD. We explore how network parameters (topology and signal control) affect two key characteristics of an MFD, (i) the network capacity and (ii) the density range for which the network capacity is maximum. We first provide an analytical proof that simplifies the estimation of the density range for which the network capacity is maximum by utilizing only spatial and control parameters of the network. We also investigate how sensitive are these two characteristics in small changes of the parameters. Afterwards, we relax the deterministic character of the parameters and investigate how variations in the signal offsets and the link lengths affect network capacity and density range. These results can be utilized to develop efficient control strategies for a series of signalized intersections as these variations can describe not only differences in network parameters, but also different characteristics in driver behavior. Later, we imitate the effect of incoming turns in a long arterial and we show that these turns can significantly decrease the network capacity even if vehicle density remains unchanged. To precisely describe all the above phenomena we initially provide some analytical proofs for a simplification of the variational theory approach and then we develop a simulator to study the non-deterministic effects.

2. A note on variational theory

Daganzo and Geroliminis (2008) developed a moving observer method to show that the average flow-density states of any urban street without turning movements must be bounded from above by a concave curve. The section also shows that, under the assumptions of variational theory, this curve is the locus of the possible (steady) traffic states for the street; i.e., it is its MFD.

Their method builds on a recent finding of Daganzo (2005), which showed that kinematic wave theory traffic problems with a concave flow-density relation are shortest (least cost) path problems. Thus, the centerpiece of variational theory (as is the fundamental diagram for kinematic wave theory) is a relative capacity ("cost") function (CF), $r(u)$, that describes each homogeneous portion of the street. This function is related to the known fundamental diagram (FD) of kinematic wave theory Q. Physically, the CF gives the maximum rate at which vehicles can pass an observer moving with speed u and not interacting with traffic; i.e., the street's capacity from the observer's frame of reference. Linear CFs correspond to triangular FDs. Daganzo (2005) assumed a linear CF characterized by the following parameters: k_0 (optimal density), u_f (free flow speed), κ (jam density), w (backward wave speed), s (capacity), and r (maximum passing rate). CF line crosses points $(u_f, 0)$, $(0, s)$, $(-w, r)$ and has a slope equal to $-k_0$. Other applications of variational theory in modeling traffic phenomena can be found in Laval and Leclercq (2008) and Daganzo and Menendez (2005).

A second element of variational theory is the set of "valid" observer paths on the (t, x) plane starting from arbitrary points on the boundary at $t = 0$ and ending at a later time, $t_0 > 0$. A path is "valid" if the observer's average speed in any time interval is in the range $[-w, u_f]$. If $\mathcal{P}$ is one such path, $u_\mathcal{P}$ be the average speed for the complete path, and $\Delta(\mathcal{P})$ is the path's cost which is evaluated with $r(u)$, $\Delta(\mathcal{P})$ bounds from above the change in vehicle number that could possibly be seen by observer $\mathcal{P}$. Thus, the quantity:

$$R(u) = \lim_{t_0 \to \infty} \inf_{\mathcal{P}} \{\Delta(\mathcal{P}) : u_\mathcal{P} = u\}/t_o \tag{1}$$

is an upper bound to the average rate at which traffic can overtake any observer that travels with average speed *u* for a long time. Note that (1) is a shortest path problem, and that $R(0)$ is the system capacity. Building on Eq. (1), Daganzo and Geroliminis (2008) proved that a ring's MFD with periodic characteristics in time and space (traffic signals every *L* meters with common green duration *G* and cycle *C* and no turns), $Q = Q(k)$, is concave and given by Eq. (2), i.e., it is not an upper bound, but a tight relation. Fig. 1a illustrates that Eq. (2) is the lower envelope of the one-parameter family of lines on the $k - q$ plane defined by $q = ku + R(u)$ with *u* as the parameter. Note this equation also describes the passing rate of an observer moving with constant speed *u* in a stationary traffic stream with flow *q* and density *k*. The main difference is that traffic signals create non-stationary conditions as vehicles stop at traffic signals and this relation does not apply in all cases. We call these lines "cuts" because they individually impose constraints of the form: $q \leqslant ku + R(u)$ on the macroscopic flow-density pairs that are feasible on a homogeneous street.

$$q = \inf_u \{ku + R(u)\} \tag{2}$$

Because evaluating $R(u)$ in Eq. (2) for all *u* can be tedious, Daganzo and Geroliminis (2008) proposed instead using three families of "practical cuts" that jointly bound the MFD from above, albeit not tightly. It has been shown (Daganzo, 2005) that for linear CF's, an optimal path always exists that is piecewise linear: either following an intersection line or else slanting up or down with slope u_f or $-w$. The practical cuts are based on observers that can move with only 3 speeds: u_f, 0 or *w* and stop at intersections during red times and possibly during green periods as well. Recall that an observer's cost rate (maximum passing rate) is $q_B(t)$ if the observer is standing at intersection with capacity $q_B(t) \leqslant s$ and otherwise it is given by a linear CF, i.e. it is either 0, *s* or *r* depending on the observer's speed.

Fig. 1b and c provide an explanation how the "practical cuts" are estimated for a series of intersections with common length and signal settings (green *G*, cycle *C* and offset δ). Offset is the time difference between the starting of the green phase
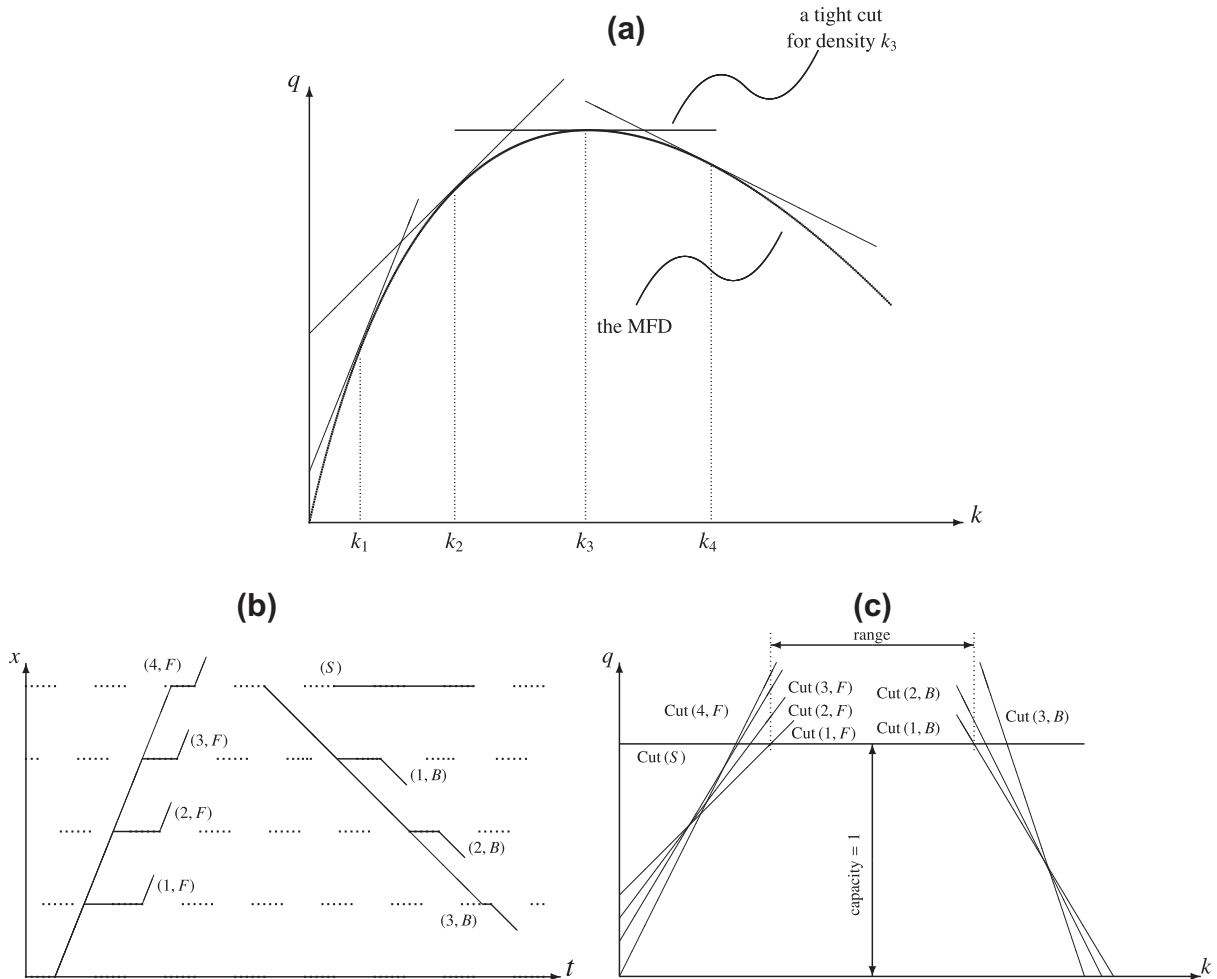

Fig. 1. (a) The MFD defined by a 1-parameter family of "cuts" (Daganzo and Geroliminis, 2008) and both forward, backward and stationary observers in a time–space (b) and their associated "cuts" in a network flow-density diagram (c).

for two consecutive signals. If the green phase of the signal in the downstream starts later than the green phase of the upstream then δ is positive for the downstream traffic signal and vice versa. Case $(4,F)$ shows the fastest moving observer, who runs with speed u_f and stops only at red phases once every 4 signals. No vehicles are passing him, the first "cut" crosses the $k - q$ plane at $(0,0)$ and has a slope equal to the average speed of the observer. The 2nd observer $(3,F)$ still runs at u_f but stops during the green period every third signal. Thus, he has a smaller average speed and vehicles are passing him at rate s when waiting in green phases. This passing rate is shown in the second "cut" as the constant of the line that crosses axis q at $q > 0$. The 3rd observer $(2,F)$ stops in every two signals whereas the fourth $(1,F)$ stops in every signal, while the fifth observer named as (S) is the stationary observer with zero speed. Cases $(3,B)$–$(1,B)$ show the backward moving observers, who are passed at rate r when moving in the opposite direction and at rate s when waiting in green phases. The lower envelope of all these "cuts" produces the MFD.

The authors conjectured that the following regularity conditions should ensure a good analytical approximation of the MFD: (i) a steady and distributed demand; (ii) a redundant network ensuring that drivers have many route choices and that most links are on many desirable routes; (iii) a homogeneous network with similar links; and (iv) links with an approximate FD that is not significantly affected by turning movements when flow is steady. Conditions (i)–(iii) should create a near-equilibrium as in Wardrop (1952) with similar travel times on all links; and, since the links are similar, with similar densities too. Condition (iv) implies that the variational theory method, applied to a single link with many efficient cuts, yields a tight MFD. The estimated MFD (despite its simplistic approach) fits well empirical and simulated data for Yokohama and San Francisco, two networks that only roughly meet the regularity conditions.

In this work, we focus on two important parameters of the MFD, produced by the family of cuts. The first one is q_{max}, the maximum value of the dimensionless network capacity, normalized by the maximum capacity of a single link sG/C, where G is the green duration during a cycle C. The second one is r_d, the range of density for which this maximum value occurs. A negative range does not have any physical meaning, except that capacity is less than 1. A large value of positive range can be considered as a metric of robustness for the system, which can retain its maximum capacity values for multiple congestion levels. We investigate these two parameters for different types of networks with homogeneous or heterogeneous characteristics of topology and signal settings.

3. Homogeneous networks

The methodology of moving observers in Section 2 could have a potential application in real cities in cases data from loop detectors are not available or do not cover the whole region. Nevertheless, the large number of cuts, would make the application not straightforward. In this section we provide an analytical proof, which significantly decreases the number of required cuts for the estimation of the range r_d. We prove that in case of homogeneous networks, the slowest forward and backward and the stationary observers suffice to estimate the same value of r_d, by running all observers described in Daganzo and Geroliminis (2008). This proof provides us the ability to obtain closed form analytical equations for r_d. We start with Lemmas 1 and 2, which prove some mathematical operations, necessary for the next steps. In Lemmas 3 (forward) and 4 (backward) we prove that the slowest moving observers provide tighter cuts than all the non-fastest observers. Lemmas 5 and 6 complete the proof by comparing with the fastest observers, who run at u_f or w and stop only at real red phases. In the end of the section, Corollary 1 provides the closed form of analytical equations of r_d for different cases. The reader can omit the proofs without lack of continuity for the rest of the paper.

In Daganzo and Geroliminis (2008), for an observer who stops every γ signals for extended red phase, delay at each stop d_γ and average speed u_γ are defined as:

$$d_\gamma = C \left(\left\lceil \frac{\gamma L/u_f - \delta}{C} \right\rceil - \frac{\gamma L/u_f - \delta}{C} \right) \tag{3}$$

$$u_\gamma = \frac{\gamma L}{d_\gamma + \gamma L/u_f}. \tag{4}$$

$\gamma_{\max}$ can be calculated as:

$$\gamma_{\max} = 1 + \max\{\gamma : \gamma(L/u_f - \delta)/C - \lfloor \gamma(L/u_f - \delta)/C \rfloor \leqslant G/C\}. \tag{5}$$

The time spent in extended red phase, f_γ equals to:

$$f_\gamma = \frac{d_\gamma - C + G}{\gamma L/u_f + d_\gamma} \text{ for } \gamma = 1, 2, \ldots, \gamma_{\max} - 1. \tag{6}$$

Similar formulations can be applied for backward moving observer, by changing u_f with w and δ with $\delta_w = C - \delta$. To differentiate them from the values given for forward moving observers, we define the same values for backward moving observers with $d_\gamma^b, u_\gamma^b, \gamma_{\max}^b$ and f_γ^b respectively.

In these set of lemmas and proofs, we want to show that, slowest forward and backward moving observers' cuts always have the tightest cut on stationary moving observers cut. We define, Cut(γ,F) and Cut(γ,B) as forward and backward moving observers' cuts which stops in an extended red or red (if $\gamma = \gamma_{\max}$ or $\gamma_{\max}^b$) phase every γ traffic signal. Cut(S) is the cut of

stationary moving observer. In order to facilitate proofs' readability, let us introduce two types of dimensionless variables Φ_γ and Φ_γ^b for forward and backward moving observers such that:

$$\Phi_\gamma = \left\lceil \frac{\gamma L/u_f - \delta}{C} \right\rceil \text{ and } \Phi_\gamma^b = \left\lceil \frac{\gamma L/w - \delta_w}{C} \right\rceil. \tag{7}$$

With the help of these variables, one can find k-coordinates of the intersection points of moving observers' cuts with stationary observers cut as follows:

$$\{\text{Cut}(\gamma, F) \cap \text{Cut}(S)\}_k = s\left[\frac{C-G}{\gamma L}(1 - \Phi_\gamma) + \frac{1}{u_f} + (C - G)\left(-\frac{\delta}{CL}\right) \right] \tag{8}$$

$$\{\text{Cut}(\gamma_{\max}, F) \cap \text{Cut}(S)\}_k = s\left[\frac{G}{\gamma_{\max}L} \Phi_{\gamma_{\max}} + \frac{G\delta}{CL} \right] \tag{9}$$

$$\{\text{Cut}(\gamma, B)\, \text{Cut}(S)\}_k = s\left[\frac{C-G}{\gamma L}\left(\Phi_\gamma^b - 1\right) + \frac{C-G}{L}\left(1 - \frac{\delta}{C}\right) + \frac{1}{u_f} \right] \tag{10}$$

$$\{\text{Cut}(\gamma_{\max}^b, B) \cap \text{Cut}(S)\}_k = s\left[-\frac{G}{\gamma_{\max}^b L} \Phi_{\gamma_{\max}} - \frac{G}{CL}(C - \delta) + \frac{1}{w} + \frac{1}{u_f} \right] \tag{11}$$

Lemma 1. *For $f(x) = \frac{1}{x}(\lceil tx \rceil - 1)$ where $t \in \mathbb{R}, f(1) \leqslant f(x)$ for $x \in \mathbb{Z}^+$.*

Proof. Assume $n < t \leqslant n + 1$ for $n \in \mathbb{Z}$. Note that, this assumption includes all real values of t. From this assumption:

$$n < t \leqslant n + 1 \Rightarrow nx < tx \leqslant nx + x \Rightarrow \lceil tx \rceil \leqslant nx + x \Rightarrow f(x) = \frac{1}{x}(\lceil tx \rceil - 1) \leqslant n + 1 - \frac{1}{x}. \tag{12}$$

If we calculate $f(1)$

$$n < t \leqslant n + 1 \Rightarrow \lceil t \rceil = n + 1 \Rightarrow f(1) = n, \tag{13}$$

which means $f(1) \leqslant f(x)$ for $x \in \mathbb{Z}^+$. $\quad\square$

Lemma 2. $\frac{\gamma_{\max}(L/u_f - \delta)}{C} \notin \mathbb{Z}$ *and* $\frac{\gamma_{\max}^b(L/w - \delta_w)}{C} \notin \mathbb{Z}$.

Proof. Assume $\frac{\gamma_{\max}(L/u_f - \delta)}{C} \in \mathbb{Z}$. Then $\frac{\gamma_{\max}(L/u_f - \delta)}{C} - \left\lfloor \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rfloor > G/C$ from Inequality (5). But this is a contradiction since $\frac{\gamma_{\max}(L/u_f - \delta)}{C} \in \mathbb{Z}, \frac{\gamma_{\max}(L/u_f - \delta)}{C} = \left\lfloor \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rfloor$ and Inequality (5) does not hold. So $\frac{\gamma_{\max}(L/u_f - \delta)}{C} \notin \mathbb{Z}$. Similar proof can be applied for the second part of the Lemma by changing w with u_f and δ_w with δ. $\quad\square$

Lemma 3. *The slowest forward moving observer is the observer who has the tightest cut on stationary observers cut among all the other non-fastest forward moving observers.*

Proof. Let us define k_γ as the k-coordinate of the $\text{Cut}(\gamma, F) \cap \text{Cut}(S)$ which is given in Eq. (8). We are looking for the value of γ that maximizes k_γ, i.e. this cut will intersect with the stationary observer at the rightest possible point, creating the tightest cut. Then our problem can be defined as:

$$\arg\max_\gamma\{k_\gamma\} = \arg\max_\gamma\left\{ s\left[\frac{C-G}{\gamma L}(1 - \Phi_\gamma) + \frac{1}{u_f} + (C - G)\left(-\frac{\delta}{CL}\right) \right] \right\}. \tag{14}$$

Since the last two elements of the RHS of Eq. (14) is not a function of gamma and $s\frac{C-G}{L}$ is always positive, we can reduce the problem to:

$$\arg\max_\gamma\{k_\gamma\} = \arg\max_\gamma\left\{ \frac{1}{\gamma}(1 - \Phi_\gamma) \right\} = \arg\max_\gamma\left\{ \frac{1}{\gamma}\left[1 - \left\lceil \frac{\gamma(L/u_f - \delta)}{C} \right\rceil \right] \right\}. \tag{15}$$

Let $n = \frac{L/u_f - \delta}{C}$, then $\arg\max_\gamma\{x_\gamma\} = \arg\max_\gamma\left\{ \frac{1}{\gamma}(1 - \lceil n\gamma \rceil) \right\} = \arg\min_\gamma\left\{ \frac{1}{\gamma}(\lceil n\gamma \rceil - 1) \right\}$ \hfill (16)

From Lemma 1, $\gamma = 1$ is the solution. $\quad\square$

Lemma 4. *The slowest backward moving observer is the observer who has the tightest cut on stationary observers cut among all the other non-fastest backward moving observers.*

Proof. Let us define k_γ^b as the x-coordinate of the $\text{Cut}(\gamma, B) \cap \text{Cut}(S)$, which is given in Eq. (10). Now the tightest cut is the one that minimizes the value of k_γ^b, as the backward moving observers' cuts have a negative slope in the $k - q$ plane. Then our problem can be defined as:

$$\arg\min_\gamma \{k_\gamma^b\} = \arg\min_\gamma \left\{ s\left[\frac{C-G}{\gamma L}\left(\Phi_\gamma^b - 1 \right) + \frac{C-G}{L}\left(1 - \frac{\delta}{C} \right) + \frac{1}{u_f} \right] \right\} \tag{17}$$

$$= \arg\min_\gamma \left\{ \frac{1}{\gamma}\left[\left\lceil \frac{\gamma(L/w - \delta_w)}{C} \right\rceil - 1 \right] \right\} \tag{18}$$

since RHS of the function is not a function of γ and $s\frac{C-G}{L}$ is always positive both are removed.

Let $n = \dfrac{L/w - C + \delta}{C}$, then $\arg\min_\gamma \{k_\gamma^b\} = \arg\min_\gamma \left\{ \frac{1}{\gamma}\left(\lceil n\gamma \rceil - 1 \right) \right\}$ \hfill (19)

From Lemma 1, $\gamma = 1$ is the solution. □

Lemma 5. *The slowest forward moving observer is the observer who has the tightest cut on stationary observers cut among all the other forward moving observers.*

Proof. In Lemma 3 it is proven that, slowest forward moving observer has the tightest cut among all the other non-fastest forward moving observers. So it is enough to show $\epsilon \geqslant 0$ where ϵ is defined as:

$$\epsilon = \{\text{Cut}(1, F) \cap \text{Cut}(S)\}_k - \{\text{Cut}(\gamma_{\max}, F) \cap \text{Cut}(S)\}_k \tag{20}$$

$$= s\left[\frac{C-G}{L}(1 - \Phi_{\gamma=1}) + \frac{1}{u_f} - \frac{\delta}{L} - \frac{G}{\gamma_{\max}L}\Phi_{\gamma_{\max}} \right]. \tag{21}$$

Let $n \in \mathbb{Z}$ that satisfies

$$n - 1 < \frac{L/u_f - \delta}{C} \leqslant n \Rightarrow \left\lceil \frac{(L/u_f - \delta)}{C} \right\rceil = n, \tag{22}$$

$$\Rightarrow \epsilon = s\left[\frac{C-G}{L}(1 - n) + \frac{1}{u_f} - \frac{\delta}{L} - \frac{G}{\gamma_{\max}L}\left\lceil \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rceil \right]. \tag{23}$$

Since $\gamma_{\max}$ is the smallest value that does not satisfy Inequality (5), we have:

$$\frac{\gamma_{\max}(L/u_f - \delta)}{C} - \left\lfloor \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rfloor > \frac{G}{C} \Rightarrow \left\lfloor \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rfloor < \frac{\gamma_{\max}(L/u_f - \delta)}{C} - \frac{G}{C}. \tag{24}$$

By Lemma 2 and Inequality (24) we can write:

$$\left\lceil \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rceil - 1 = \left\lfloor \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rfloor < \frac{\gamma_{\max}(L/u_f - \delta)}{C} - \frac{G}{C}. \tag{25}$$

By combining Inequality (25) with Eq. (23):

$$\epsilon > \left[s\left(1 - \frac{G}{C} \right) \right]\left[\frac{1}{u_f} - \frac{\delta}{L} - \frac{G}{\gamma_{\max}L} + (1 - n)\frac{C}{L} \right]. \tag{26}$$

We know the first multiplicative in the RHS of Inequality (26) is positive. It is sufficient to show the second one is positive too. If we multiply both sides of Inequality (22) by γ_{max} and apply floor function on both sides:

$$\left\lfloor \frac{\gamma_{\max}(L/u_f - \delta)}{C} \right\rfloor \geqslant \gamma_{\max}(n - 1). \tag{27}$$

If Inequality (25) is combined with Inequality (27) then,

$$\frac{\gamma_{\max}(L/u_f - \delta)}{C} - \frac{G}{C} - \gamma_{\max}(n - 1) > 0. \tag{28}$$

After multiplying both sides of Inequality (28) by a positive number $\frac{C}{\gamma_{\max}L}$ we will have:

$$\frac{1}{u_f} - \frac{\delta}{L} - \frac{G}{\gamma_{\max}L} + (1 - n)\frac{C}{L} > 0, \tag{29}$$

which is the second multiplicative on the RHS of Inequality (26). Since $\epsilon > 0$, the slowest forward moving observer always has the tightest cut on stationary observer. □

Lemma 6. *The slowest backward moving observer is the observer who has the tightest cut on stationary observers cut among all the other backward moving observers.*

Proof. We will use a method similar to proof of Lemma 5. In Lemma 4, it is proven that, the slowest backward moving observer has the tightest cut among all the other non-fastest backward moving observers. So it is enough to show $\epsilon \geqslant 0$ where ϵ is defined as:

$$\epsilon = \{\text{Cut}(\gamma^b_{\max}, B) \cap \text{Cut}(S)\}_k - \{\text{Cut}(1, B) \cap \text{Cut}(S)\}_k \tag{30}$$

$$= s\left(-\frac{G}{\gamma^b_{\max}L}\Phi^b_{\gamma=1} - \frac{G}{L} + \frac{1}{w} + \frac{\delta}{L} - \frac{C-G}{L}\Phi^b_{\gamma^b_{\max}}\right). \tag{31}$$

Let $n \in \mathbb{Z}$ that satisfies

$$n - 1 < \frac{L/w - C + \delta}{C} \leqslant n \Rightarrow \left\lceil\frac{L/w - \delta_w}{C}\right\rceil = n, \tag{32}$$

$$\Rightarrow \epsilon = s\left[-\frac{G}{\gamma^b_{\max}L}\Phi^b_{\gamma^b_{\max}} - \frac{G}{L} + \frac{1}{w} + \frac{\delta}{L} - \frac{Cn}{L} + \frac{Gn}{L}\right]. \tag{33}$$

Since $\gamma^b_{\max}$ is the smallest value that does not satisfy Inequality (5), we get:

$$\left\lfloor\frac{\gamma^b_{\max}(L/w - \delta_w)}{C}\right\rfloor < \frac{\gamma^b_{\max}(L/w - \delta_w)}{C} - \frac{G}{C}. \tag{34}$$

By combining Lemma 2 with Inequality (34) we can write:

$$\left\lceil\frac{\gamma^b_{\max}(L/w - \delta_w)}{C}\right\rceil - 1 = \left\lfloor\frac{\gamma^b_{\max}(L/w - \delta_w)}{C}\right\rfloor < \frac{\gamma^b_{\max}(L/w - \delta_w)}{C} - \frac{G}{C}. \tag{35}$$

If Inequality (35) is combined with Eq. (33)

$$\epsilon > \left[s\left(1 - \frac{G}{C}\right)\right]\left(\frac{1}{w} + \frac{\delta}{L} - \frac{Cn}{L} - \frac{G}{\gamma^b_{\max}L}\right). \tag{36}$$

We know the first multiplicative in the RHS of Inequality (36) is positive. It is sufficient to show the second one is also positive. If we multiply both sides of Inequality (32) by γ^b_{max} and apply floor function on both sides:

$$\left\lfloor\frac{\gamma^b_{\max}(L/w - \delta_w)}{C}\right\rfloor \geqslant \gamma^b_{\max}(n-1). \tag{37}$$

If Inequality (35) is combined with Inequality (37) then,

$$\frac{\gamma^b_{\max}L}{wC} - \frac{\gamma^b_{\max}(C-\delta)}{C} - \frac{G}{C} - \gamma^b_{\max}(n-1) > 0. \tag{38}$$

After multiplying both sides of this inequality by a positive number $\frac{C}{\gamma^b_{\max}L}$ we will have:

$$\frac{1}{w} - \frac{C}{L} + \frac{\delta}{L} - \frac{G}{\gamma^b_{\max}L} - (n-1)\frac{C}{L} > 0, \tag{39}$$

which is the second multiplicative of Inequality (36). Since $\epsilon > 0$, the slowest backward moving observer always has the tightest cut on stationary observer. □
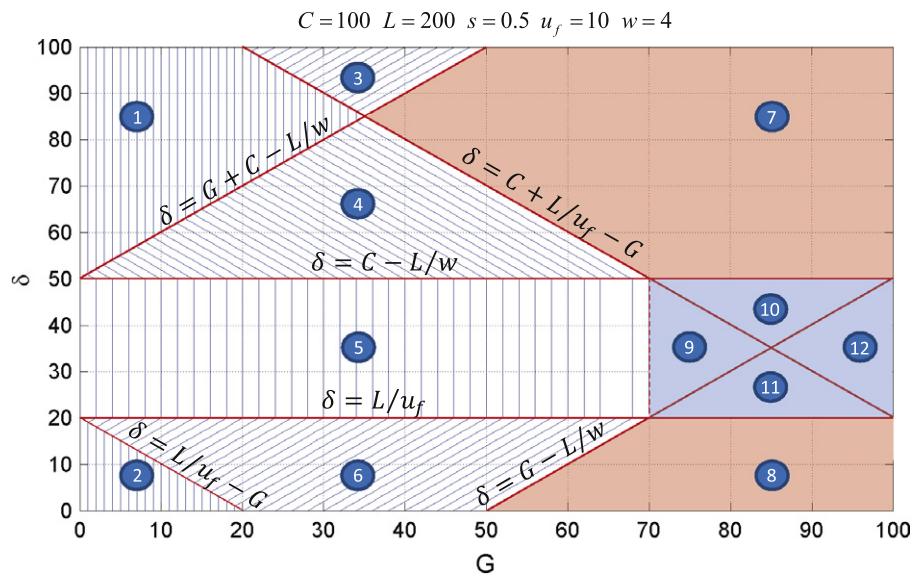
Corollary 1. *The density range of maximum capacity can be calculated by using slowest forward moving, slowest backward moving and stationary observers. For four different cases, the range can be formulated as:*

$$\begin{array}{c}\text{many forward} \\ \text{many backward}\end{array} : \quad Range_{M-M} = s\frac{C-G}{L}\left(\left\lceil\frac{L/w - \delta_w}{C}\right\rceil + \left\lceil\frac{L/u_f - \delta}{C}\right\rceil - 1\right) \tag{40}$$

$$\begin{array}{c}\text{one forward} \\ \text{one backward}\end{array} : \quad Range_{1-1} = s\left(\frac{1}{w} + \frac{1}{u_f}\right) - s\frac{G}{L}\left(\left\lceil\frac{L/w - \delta_w}{C}\right\rceil + \left\lceil\frac{L/u_f - \delta}{C}\right\rceil + 1\right) \tag{41}$$

$$\begin{array}{c}\text{many forward} \\ \text{one backward}\end{array} : \quad Range_{M-1} = s\left(\frac{C-G}{L}\left\lceil\frac{L/u_f - \delta}{C}\right\rceil - \frac{G}{L}\left\lceil\frac{L/w - \delta_w}{C}\right\rceil - \frac{C-\delta}{L} + \frac{1}{w}\right) \tag{42}$$

$$\begin{array}{c}\text{one forward} \\ \text{many backward}\end{array} : \quad Range_{1-M} = s\left(\frac{C-G}{L}\left\lceil\frac{L/w - \delta_w}{C}\right\rceil - \frac{G}{L}\left\lceil\frac{L/u_f - \delta}{C}\right\rceil - \frac{\delta}{L} + \frac{1}{u_f}\right) \tag{43}$$

$$C = 100 \quad L = 200 \quad s = 0.5 \quad u_f = 10 \quad w = 4$$

#	observers	range	#	observers	range
1	1 F - 1 B	$s\left(\frac{1}{w} + \frac{1}{u_f} - \frac{2G}{L}\right)$	7	many F - many B	0
2	1 F - 1 B	$s\left(\frac{1}{w} + \frac{1}{u_f} - \frac{2G}{L}\right)$	8	many F - many B	0
3	many F - 1 B	$s\left(\frac{1}{w} + \frac{G-\delta}{L}\right)$	9	1 F - 1 B	< 0
4	1 F - many B	$s\left(\frac{1}{u_f} + \frac{C-G-\delta}{L}\right)$	10	many F - 1 B	< 0
5	1 F - 1 B	$s\left(\frac{1}{w} + \frac{1}{u_f} - \frac{G}{L}\right)$	11	1 F - many B	< 0
6	many F - 1 B	$s\left(\frac{1}{w} + \frac{\delta-G}{L}\right)$	12	many F - many B	< 0

Fig. 2. Regions and formulations of each region according to Corollary (1) for $L/u_f + L/w < C$.

Eqs. (40)–(43) provide a closed form description of range. In case range has a negative value, this means that the network capacity is smaller than 1, i.e. smaller than sG/C. Fig. 2 provides an illustration of the estimated range for C = 100 s and L = 200 m and analytical formulae for all regions. Note that the highlighted area with blue for G between 70 and 100 s shows (δ, G) pairs with negative range (capacity strictly less than one) and the highlighted area with red shows (δ, G) pairs with zero capacity (i.e. the tightest forward and backward moving observers' cuts intersect at the same point on the stationary observers' cut). Even if there is not an analytical proof for the required cuts to estimate the value of capacity when range is negative, not all cuts are necessary. We run numerous simulations for different values of C, L, δ and G/C and we find out that by utilizing only the slowest and fastest, forward and backward moving observers (in total 4), the error in the estimation of capacity when range is negative is small. A few more observers are needed in case of small L/C values. A description of cases with smaller capacity is provided in the following sections.

4. Simulation framework

In the previous section we assumed a homogeneous network with deterministic values of all parameters (link lengths, green durations, offsets) and no turns. However, real life networks contain some variability in the network parameters and also drivers' decisions contain stochasticity. By introducing a degree of variability, analytical solutions are not anymore obtainable. Thus, we develop a simulation platform to estimate the passing rates and average speeds of forward and backward moving observers running a series of many intersections with variable characteristics. By analyzing the results of the simulation, we can identify the effect of heterogeneity in the topology and signal settings at the network capacity and density range.

While variational theory allows for changes in the network parameters, it does not give the ability to introduce drivers with different characteristics (free flow speed, capacity headway, etc.) and turning movements. But, we can mimic the effect of driver stochasticity and incoming turns by adjusting offsets and green durations. For example, consider an arterial's signal plan, which has been designed for a perfect progression, a "green wave", with offsets equal to L/u_f. By introducing some randomness in the offset, e.g. $L/u_f \pm 5$ s we can imitate the variability in the free flow travel time of the first vehicles in the platoon. Also, we will show how incoming turns affect the network capacity in cases that network density remains unchanged.

The simulation platform includes a time–space diagram with many links (∼1000) to produce more rigorous results. The network parameters (lengths, offsets and greens) are specified from the user in the beginning of the simulation in column or matrix forms. After creating the simulation environment, we initiate different types of observers at the start time of a green phase running at the free flow speed from the upstream in the direction of flow and with the backward wave speed from downstream in the direction against flow as described by the theory of Section 2. When they are moving all observers on the same direction have equal speeds (free flow or backward wave) but they have different behavior than the deterministic case at (normal or extended) red phases, as there is not a repetitive deterministic pattern due to signal settings and link lengths (e.g. we cannot say that an observer stops every three signals).

In simulation, this pattern becomes stochastic by giving probabilities that an observer will stop if he meets an extended red phase. According to Daganzo and Geroliminis (2008) an extended red phase is used to make observers stop every $1, 2, \ldots, \gamma_{max} - 1$ traffic signals. Each observer in the simulation is assigned with a probability of stopping each time he meets a green phase. Faster observers are assigned with smaller probability and slower observers with higher one. For instance, if the probability assigned to this observer is 0, this observer will only stop when he hits red whereas the observer with probability 1 will stop in every signal even though he hits green. If we consider an observer with probability p, he will pass on green phases with probability $1 - p$. For each observer we have to estimate two values, the average speed and the average passing rate. Average speed can be calculated by dividing the sum of link lengths to the total travel time. Similarly, we track the number of passing/passed vehicles for each observer during the simulation and divide them by the total travel time. For the same value of p, a number of iterations is performed (∼10) and different paths can be constructed because of the stochastic behavior of moving observers. A lower envelope of cuts is estimated to be consistent with VT. Nevertheless, we have noticed that all iterations for the same value of p and different random seeds give almost identical results. An informal pseudocode for a single forward moving observer can be seen below. The one for the backward moving observer is the same. The only difference is the speed, the start and the direction of the movement:

Note that in the pseudocode given above, green phase matrix **G** is represented as a set of unions of real number intervals $\mathbf{G}_i$ to be consistent with mathematical notations. They represent the same parameter set. Fig. 3b shows a time space diagram

---

input: n (number of traffic signals), **l** (link length vector), $\mathbf{G}_i$ (green set for signal i),
  p (extended red phase probability), u_f (free flow speed) and s (capacity)
output: v (average speed) and PR (average passing rate)

1. Set time t to first green signals start time t_0 and total passing rate TPR to 0.
2. Repeat if current traffic signal counter $i < n$
    (a) Advance the time by travel time on that link: $t \leftarrow t + \frac{L[i]}{u_f}$.
    (b) Update the current traffic signal counter $i \leftarrow i + 1$.
    (c) If the trajectory hits a green ($t \in \mathbf{G}_i$)
        i. If it is an extended red phase ($p > \mathrm{RAND}(0,1)$)
           A. Find extended red phase duration k.
           B. Update total passing rate: $TPR \leftarrow TPR + ks$.
           C. Advance time t to next green.
        ii. If it is not an extended red phase ($p \le \mathrm{RAND}(0,1)$)
           A. Do nothing.
    (d) else if hits a red ($t \notin \mathbf{G}_i$)
        i. Advance time t to next green.
3. Calculate the average speed $v \leftarrow \frac{\sum_{i=1}^{n} L_i}{t - t_0}$ and the average passing rate $PR = \frac{TPR}{t - t_0}$.
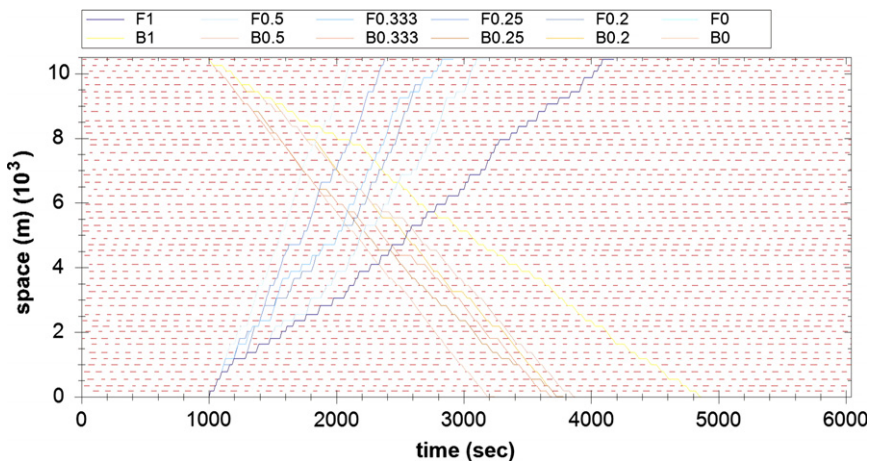

Fig. 3. Simulation platform: (a) pseudocode and (b) time–space diagram.

with the forward (F) and backward (B) moving observers with different values of stopping probabilities (value is shown on the top of the figure).

4.1. Incoming turns

Daganzo and Geroliminis (2008) showed that an MFD must arise for single-route networks with a fixed number of vehicles in circulation (i.e., periodic boundary conditions and no turns). The authors also conjecture that the MFD formulae should apply to a network of intersecting routes if the numbers of vehicles in these routes are similar and roughly constant over time. We now address the effect of incoming turns in a single-route network by introducing bottlenecks of variable capacity in the proximity of the traffic signals.

Incoming turns from cross streets can significantly decrease the performance of a signalized intersection as they (i) interrupt the progression of green waves for properly timed signals and (ii) decrease the available storage capacity of the link and can cause the occurrence of spillbacks. This problem is difficult to solve in the general case because inflow needs to be separated in two classes (upstream through and incoming turns) and the "cuts" approach cannot directly identify the mixing of the vehicle origins. We address the above phenomena by changing the signal and cost function characteristics for forward and backward moving observers. To be consistent with variational theory, these turns should not significantly change the link density from one link to another, i.e. the incoming turns are considered as local link phenomena, for example there is a similar number of outgoing turns in every link to keep density constant. We also assume that incoming turns are served first because they enter the link when the through approach is in a red phase. This might not be exact when residual queues still exist or incoming turns might occur from an unsignalized intersection or a parking lot (internal source).

Consider now a queue of incoming vehicles from cross streets, Q, which entered when the signal was red for the through movement (Fig. 4a). If these vehicles did not exist, a forward moving vehicle would follow trajectory F (Fig. 4a) and would stop for some time at the traffic signal stop line. Because of the queue of incoming turns, the upstream vehicle needs to stop between points X_1 and X_2, follow trajectory F_τ and cross the intersection τ seconds later, where $\tau = sQ$. According to variational theory the cost (passing rate) of a forward moving observer who stops in the middle of a link between X_1 and X_2 is st where t is the duration of stop. But, in reality no vehicles can overpass this observer while stopping, because in front of him there is a queue of vehicles entered from a cross street. This can be shown if one estimates the change between points X_1 and X_2 in the Moskowitz function $N(t,x)$ which expresses the cumulative number of vehicles in the $t − x$ plane. It is also known that this change is the same for all possible paths between A and B (Daganzo, 2005). Moskowitz function value
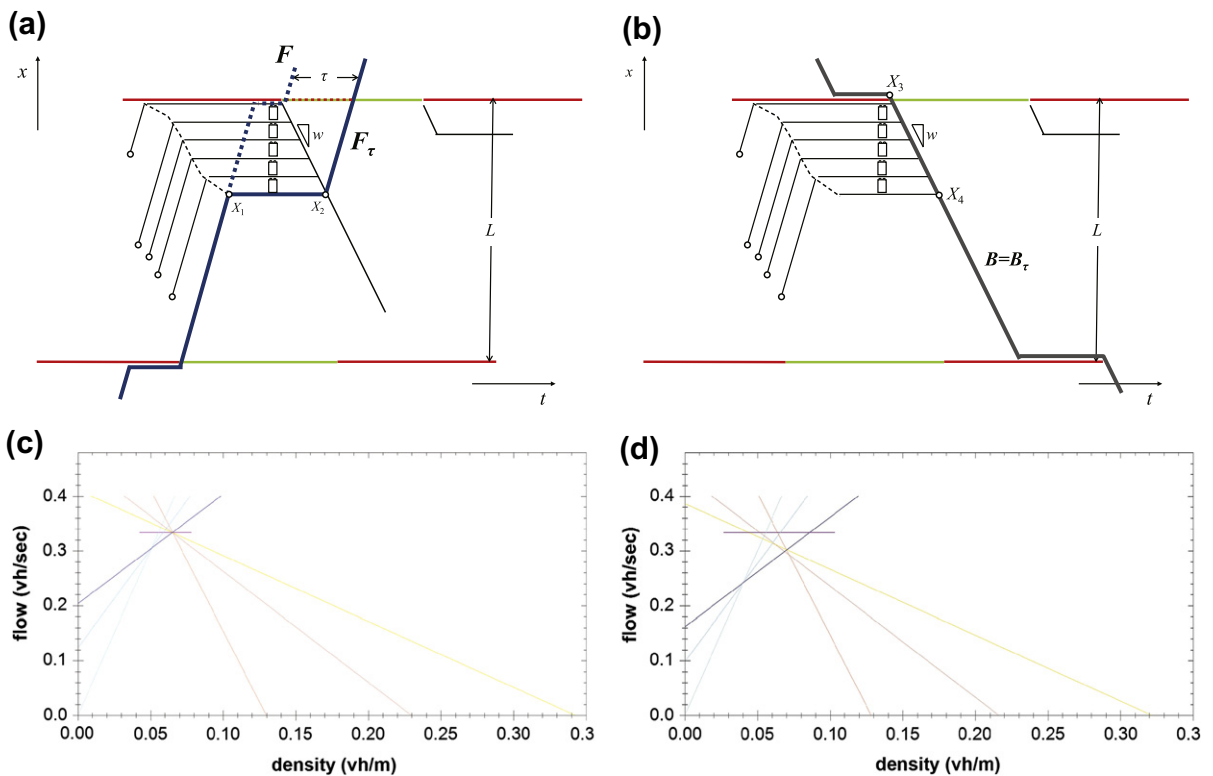

Fig. 4. Integrating the effect of incoming turns within variational theory: time–space diagrams for forward (a) and backward (b) moving observers with (**F** and **B**) and without turns (**F**$_\tau$ and **B**$_\tau$), flow-density diagrams without (c) and with (d) turns.

changes during this extended red phase by an amount τs. Thus, we imitate this effect by increasing the red phase of the signal by τ, only for the forward moving observers. So, these observers, will follow trajectory $\pmb{F}_\tau$ instead of $\pmb{F}$. But, this extended red in the beginning of the green has passing rate zero, not s. Nevertheless, the stationary observer in front of the traffic signal continues to count vehicles for the whole duration of the real green phase as the incoming turns are served in the first τ seconds of green.

For the backward moving observer, $\pmb{B}_\tau$ (Fig. 4b) the approach is slightly different. This observer does not need to stop in the extended red phase of τ seconds. But when traveling backwards between points X_3 and X_4, its passing rate is not r, but zero. Thus, the queued vehicles from cross street, give the ability to the observer to travel in this queue with zero cost.

Based on the above, for the forward moving observers, both the speed and passing rate decreases; for the backward moving observers, only the passing rate decreases; while for the stationary observer there is no change. Thus, tighter cuts are created which can decrease both the range and the capacity of the MFD. An example is shown in Fig. 4c and d with and without turning effects. These transformations in the trajectories of the observers are in accordance with the Lagrangian variational principles, as expressed by Daganzo (2005) and Leclercq et al. (2007) and they are valid even when the Moskowitz function is not continuous and experiences step-jumps in the time–space profile (e.g. because of incoming turns in our case). Analysis of the results is provided in the next section.

In reality, there is often a spatial correlation between incoming turns, i.e. a series of successive intersections load the main route with net positive inflows from cross streets. In this case the proposed approach lacks of methodological correctness. The reason is that these turns might increase the density in the aforementioned links of the main route and result in an inhomogeneous distribution of density, while our approach considers links with similar density. This case is not addressed in the current paper, but future work can utilize our results by partitioning the route in sub-routes and estimate individual fundamental diagrams with variational theory. A research priority should be how to integrate serial or parallel routes of MFDs, with different average densities to describe the dynamics of traffic flow.

5. Results

In this part, we firstly investigate the deterministic cases which is solved by the analytical formulae given in "homogeneous networks" section and then continue with the "simulation" results. In order to generate isoquants either the analytical formulation (Corollary 1) or the simulation method (Fig. 3) is used for finite number of (x, y) values which are elements of Cartesian product of number of predefined points on x and y axis of the given graph. For example for Fig. 5a $(x, y) \in \{0.10, 0.11, \ldots, 0.90\} \times \{0, 1, \ldots, 60\}$, for Fig. 6c $(x, y) \in \{0.10, 0.11, \ldots, 0.90\} \times \{U(180, 180), U(179, 181), \ldots, U(80, 280)\}$ where $U(*)$ represents uniform distribution.

As it is defined in Fig. 1c, dimensionless capacity (values between 0 and 1) is the ratio of the maximum flow q_{max} to the flow observed by the stationary observer which equals to sG/C. Range is the difference between maximum and minimum density which yields the maximum flow q_{max}. Since the value of the negative range does not mean anything and capacity equals to 1 if the range is nonnegative, it is possible to merge contour lines for both range and capacity at the same graph. These merged graphs can be seen in Figs. 5–8. Note that, blue stands for the range whereas red stands for the capacity.

5.1. Deterministic network parameters

Fig. 5 summarizes how range and capacity change for different values of offsets, green and cycle durations and link length. Fig. 5a-c plot range and capacity with δ and G/C for three different cases (i) $C = 60$ s, $L = 110$ m (short link, small cycle), (ii) $C = 90$ s, $L = 225$ m (long link, medium cycle) and (iii) $C = 120$ s, $L = 180$ m (medium link, long cycle). The white area between the blue and red isoquants in graphs represents scenario with capacity 1 and range zero, i.e. tightest forward and backward observers intersect with the stationary one at the same point. Note that for a range of δ (e.g. $8 - 38$ s in Fig. 5a), density range is invariant with offset. Note also that by increasing G/C after some value, capacity decreases (remember that the graphs show dimensionless capacity). Note also that the effect of bad coordination in the capacity of a short link is much more significant in case of short links and when the cycle is longer (compare Fig. 5b with Fig. 5c). Notice that not only perfect coordination in offset (L/u_f) but also the values between L/u_f and $C - L/w$ gives the maximum range for any given G/C ratio. Furthermore, in this region, range is independent of the offset which was also proved in Section 3. Certainly, the positive range region is larger either when L is larger or C is smaller. This fact can be expected from the Eqs. (40)–(43).

Fig. 5d shows how combinations of green duration and green over cycle ratio (G vs. G/C) affect range and capacity for medium size links with $L = 180$ m and bad offset (first vehicle arrives 13.5 s before the beginning of green phase). Note there is a boundary line which shows under what cases there is a capacity drop. Note also that this drop occurs for a larger range of green duration when the cycle is longer. This is an intuitive observation, as longer cycles create longer queues that can spillback and decrease the output of intersections. For $C = 90$ s, by increasing G/C from 0.5 to 0.7 (40% increase) the improvement in the maximum number of vehicles that can be served is too small (10% increase), 0.25 vh/s vs. 0.28 vh/s (the values have been obtained by multiplying the numbers of the graph with sG/C). This possibly means that the additional G/C can be utilized to serve cross streets with less delays. Similar graphs can be produced for different values of offsets and link lengths.

Fig. 5e and f show the effect of the link length on range and capacity. Range is increasing as the link length increases in both graphs. Note that for the same value of offset, as increase in G causes a decrease in range, as the stationary observer has
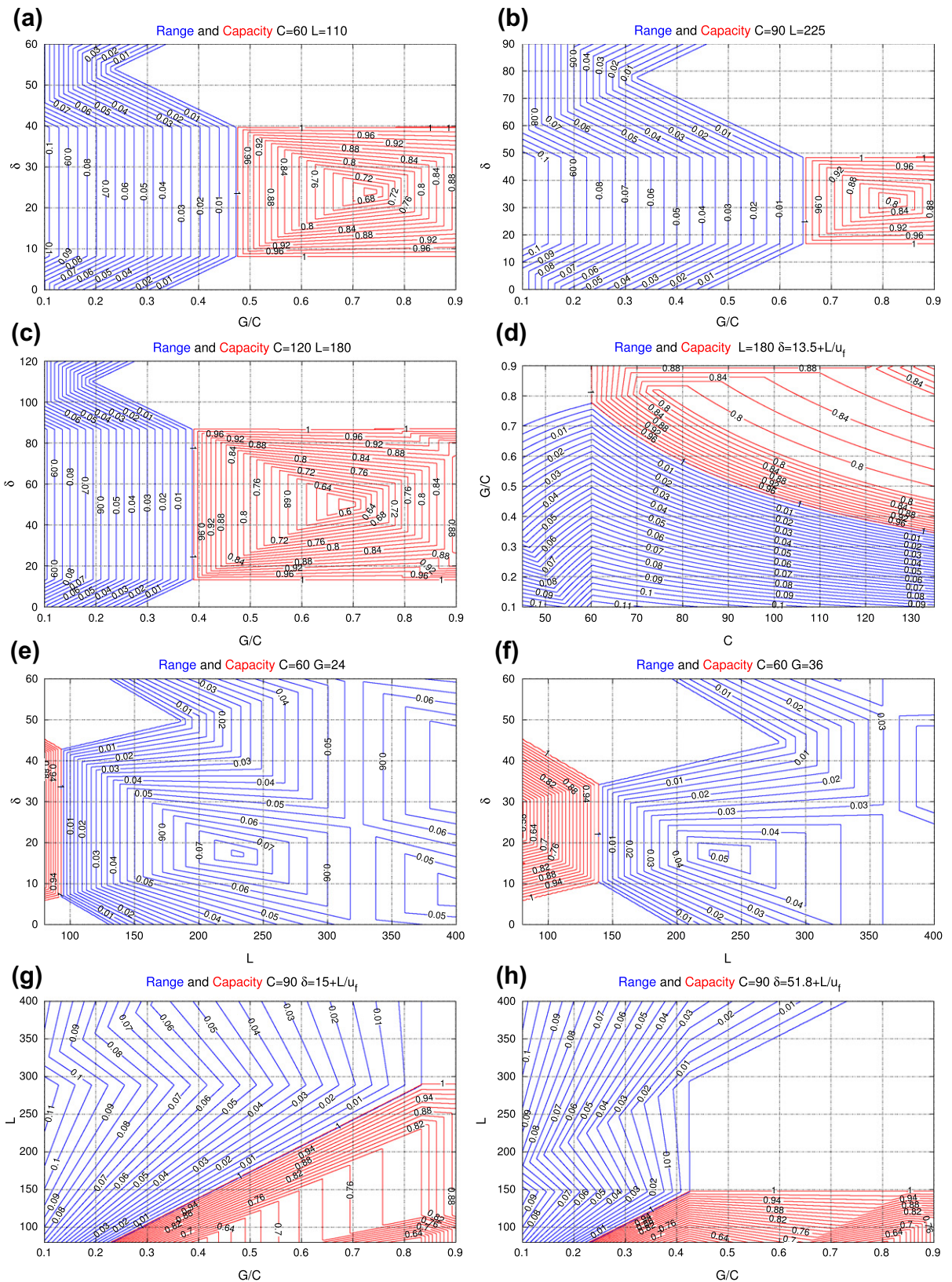
Fig. 5. Deterministic cases (homogeneous networks). Range and capacity for different values of topological and signal characteristics.
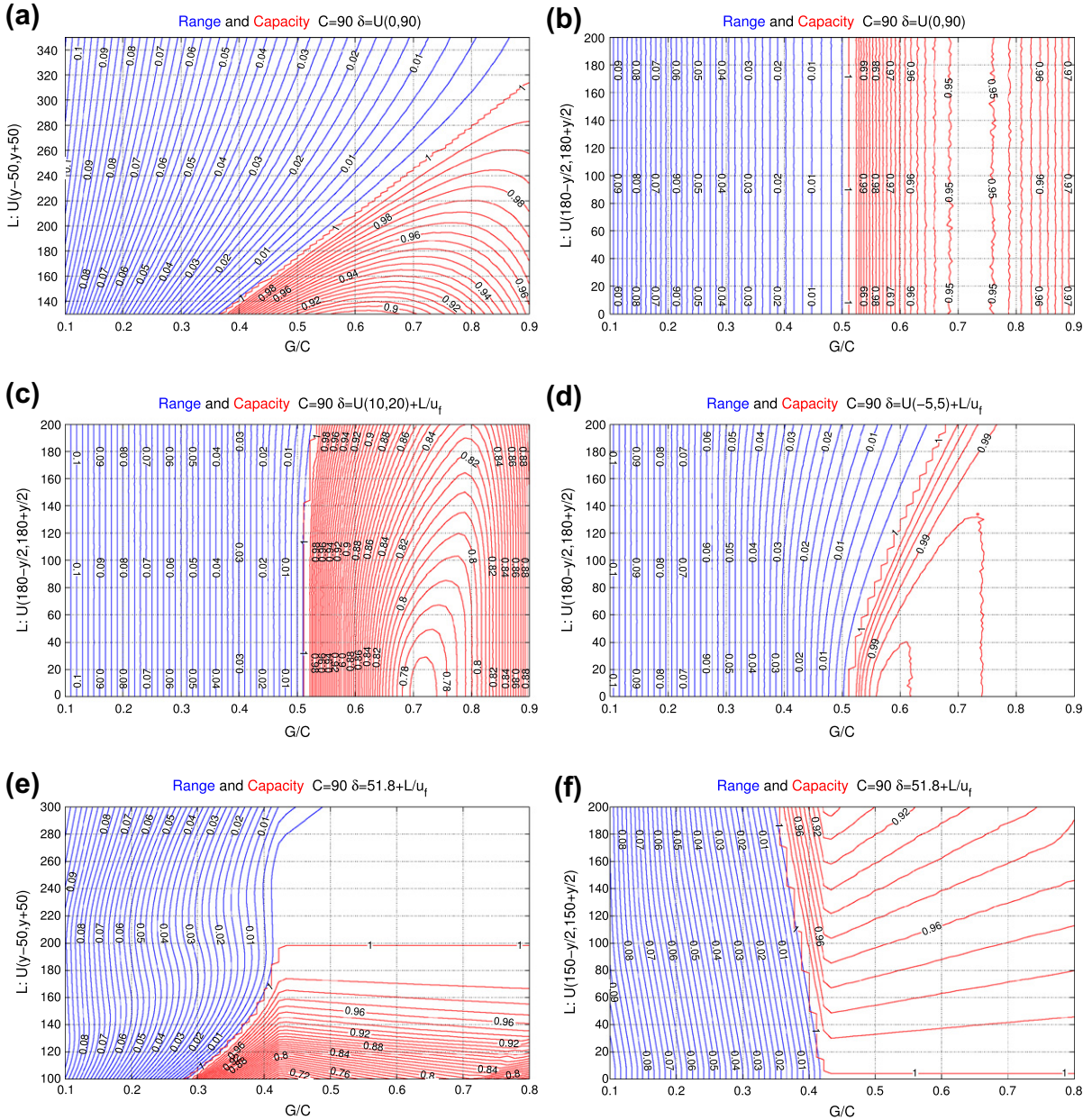
Fig. 6. Stochastic *L*.

a higher passing rate value. An interesting observation is that routes with higher green durations (mainly the ones that carry a lot of traffic) can experience smaller capacities (q_{max} < 1) for a wider range of link lengths (critical length is 80 m for *G* = 24 and 140 m for *G* = 36 s). When *L* is large, capacity is always 1 for any value of offset. In this case, we can choose offset in a way to maximize the range, as this means that the signal can operate at capacity for a wide range of densities. Also, values in the white regions might not be stable as small changes in demand can create spillbacks or capacity drops.

In Fig. 5g and h, we investigate the effect of length and green phase. Note that the boundary line for capacity less than 1, is a piecewise linear function of *L* and *G*/*C*. The locus of points that produce capacity drop is much larger set in case of bad offsets (Fig. 5g). Note also that for *δ* = 51.8 s there is a large number of pairs with capacity 1 and range zero, i.e. all tight cuts are intersecting at the same point. In these cases heterogeneity in the distribution of congestion in the network might create a significant capacity drop (e.g. Mazloumian et al., 2010). For example for *L* = 150 m and *δ* = 15 s (bad offset where vehicles have to stop in every signal), *G*/*C* ratio greater than 0.43, will cause not full utilization of signal capacity while for the case of Fig. 5h there is no capacity drop for any value of *G*/*C*. These graphs show the importance of the described methodology in
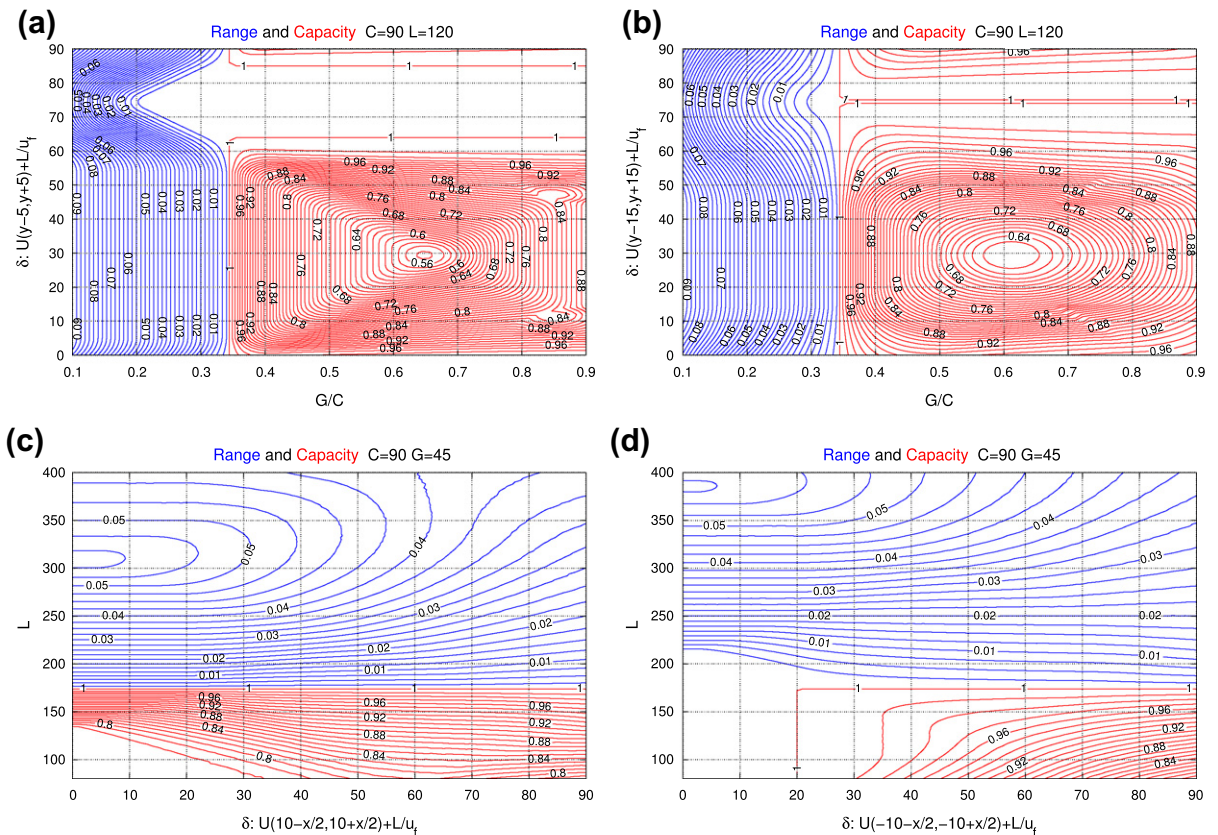
Fig. 7. Stochastic δ.

estimating the effect of signal characteristics and link length in the network capacity. All of the above analysis would not be possible using standard traffic engineering techniques (e.g. the Highway Capacity Manual).

5.2. Stochastic network parameters

We now utilize the simulation platform to identify the effect of variability when compared with the deterministic cases described before. The results presented assume a uniform distribution for offsets and link lengths, $U(min, max)$. One can apply different distributions if needed. We analyze two sets of variations for the different topological and signal parameters: (i) the mean value is constant and range of the variable changes and (ii) the range is constant and the mean of the variable changes. The next three subsections present results for variations in the effect of link lengths, offsets and incoming turns.

5.2.1. Variations in link length

In the graphs given in Fig. 6, both the change in the mean (for a given variance) and the variance (for a given mean) of the distribution of L are analyzed. In Fig. 6a and b, offsets are random at every intersection (varying uniformly between 0 and cycle length) and link length L has a uniform distribution. In Fig. 6a, L has a constant range of 100 m and variable mean, while in Fig. 6b, L has a constant mean (medium length link with $E[L] = 180$ m) and variance varies. In the first case we see that the critical link length for which a capacity drop occurs increases with the G/C ratio (almost linearly). For $G/C \leqslant 0.3$ capacity is always fully utilized, while for larger values we might observe a drop up to 10% for $G/C = 0.6$. In other words, shorter link lengths are more sensitive to green ratio. Once compared with the deterministic case of Fig. 5g (bad offset), we observe that the capacity drop is less intense in case of random offsets. Fig. 6b is intuitive to understand, as vertical isoquants mean that increase in the length variability have no effect in capacity or range for medium length links and random offsets. Thus, we are interested in identifying the critical length variability which changes the deterministic results.

In the Fig. 6c and d, the effect of G/C ratio and variance of link length is analyzed for links with average length $E[L] = 180$ m in the case of almost perfectly (first vehicle from upstream arrives in the beginning of green) and badly coordinated signals (first vehicle arrives 15 s before the end or red phase). A small variation ±5 s has been introduced for good and bad offsets. Results show that length variability increase has no effect for values of G/C smaller than 0.45 (capacity is always 1 for these values). But, when capacity is less than 1 there is a range of G/C where significant increase in capacity is observed for the case
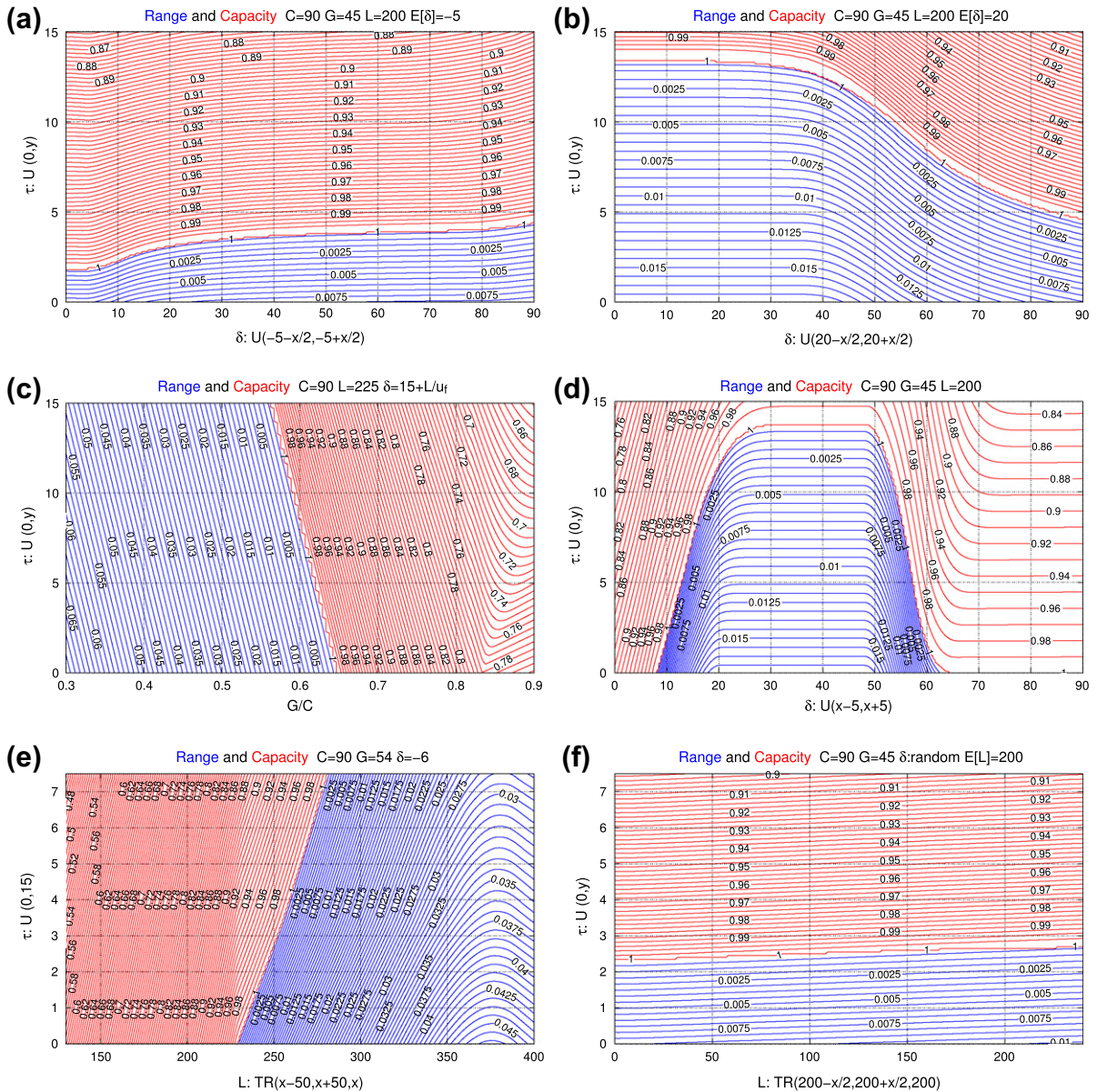
Fig. 8. The effect of incoming turns in capacity and range.

of bad offsets (about 15% change for $G/C = 0.65$ once comparing deterministic L and highly variable L between 80 and 280 m). Good offsets are not significantly affected (change not more than 2%).

Fig. 6e and f have the same structure as Fig. 6a and b but the offset is deterministic at a value of 51.8 s, and the mean length is 150 m in Fig. 6f. This is the value for which capacity is equal to 1 in the case of Fig. 5h. It is clear that for small values of G/C there is a positive range (capacity is always 1) and neither variability in length, nor change of the mean have an effect. But for $G/C \geqslant 0.4$ the situation is different. In Fig. 6e, increase of G/C does not significantly change the capacity for a given length structure (note the horizontal lines). But, as expected the longer the mean the longer the capacity.

Fig. 6f has some additional interesting characteristics. We note that when capacity is less than 1 ($G/C \geqslant 0.4$), increase in the length variability decreases the value of capacity (maximum drop about 6%). This is the opposite once compared with Fig. 6c, where in case of bad offsets, increase in length variability creates an increase in the value of the capacity. To explain this, we need to look at the deterministic graphs for the specific values of offsets (Fig. 5g-h). For example, let's focus on $G/C = 0.7$. For $L = 150$ m (Fig. 5h), an increase in the variability will result in many short links with smaller capacity, while the longer links will have same value of capacity (equal to 1). For $L = 180$ m (Fig. 5g), as variability increases, most of the short links will have capacity around 0.76 (constant) while capacity will increase for longer links. Note that this explanation

provides is a qualitative insights as the stochastic case (Figs. 6 and 7) cannot be reproduced by a linear combination of the deterministic examples.

5.2.2. Variations in offset

We first investigate the effect of small (±5 s) and large (±15 s) variations in offset by changing mean offset and G/C. The results are summarized in Fig. 7a and b. If these graphs are compared with the deterministic graphs (Fig. 5a–c), we can say that small offset variations have no significant effect both in capacity and range. This means that small differences in drivers' characteristics (e.g. free-flow speed, reaction time) cannot decrease the performance of traffic signals. However, in case of large offset variations (in case of poorly designed signals) the effect can be significant, especially in regions with capacity less than 1. The choice of appropriate offsets is more critical for routes or subnetworks which carry high demand and have large G/C values, e.g. high directional flows in the morning peak with small cross-street flows. Note that capacity can decrease by an amount of 20–30% for G/C in the range of 60–70%.

In the remaining two graphs, the range of the offset is investigated for good (the first vehicle from upstream arrives 10 s after the beginning of green) and bad coordination (the first vehicle arrives 10 s before the end of red). All the parameters except the means of the offsets are the same in (c) and (d). When capacity is smaller than 1 (for $L < 140$ m), higher offset variability improves the capacity value in case of bad offsets and has a negative effect in case of good offsets. This result is intuitive as more (less) vehicles will hit the green phase during bad (good) coordination. For long links, offset variability decreases the capacity as for a given L and G/C range is maximized when there are only one forward and one backward observers (Fig. 5a–c).

5.2.3. The effect of incoming turns

We now show that incoming turns from cross streets can significantly decrease the performance of a signalized intersection in some cases as they interrupt the progression of properly timed signals and decrease the available storage capacity of the link.

In all graphs of Fig. 8 the vertical axes is the amount of incoming turns (expressed as the extended red phase $\tau = sQ$, which is assumed uniform between 0 and an increasing value). The first two graphs on the top show the effect of turns as the variability of offsets increases for good (left graph) and bad (right graph) offsets. In case of almost perfect offsets the effect of turns is very significant because the value of range for zero turns is very small. On the other hand, the bad offsets can absorb a high number of turns without capacity decrease. Notice that the values of the two graphs coincide for $\delta = 90$ s as this represents the case of random offsets. A signal timing with bad offsets can absorb up to 13 s of turning (6.5 vehicles), while even one incoming vehicle can create problems for good coordination. But even in case of bad offsets, large G/C is problematic as range is smaller even for zero turns. Thus, in case of incoming turns, the signal plans should be chosen in a way that maximizes the range as capacity can be significantly decrease. For example in Fig. 8d one can see that a bad offset with higher range is much more robust than a good offset with small range. Of course if signals are undersaturated, they will operate in values much less than capacity and the effect of turns will be minimal. But, in this paper we mainly investigate signal performance in high demand conditions. Also, from Fig. 8e and f it is clear that as length increases the effect of incoming turns becomes smaller because it is more difficult to have queue spillbacks.

6. Conclusions

In this paper we have provided several extensions and refinements in the variational theory of traffic flow, which provides analytical formulae for the macroscopic fundamental diagram of urban networks. In our study we investigated the effect that have in the MFD, different degrees of variability in link lengths and signal characteristics for different city topologies and signal structures. We have integrated the effect of incoming turns in the estimation of the MFD and we showed that in many cases network capacity can significantly decrease. The scalability of flows from a series of links to large traffic networks is not a straightforward transformation. Route or network capacity can be significantly smaller than the capacity of a single link, because of the correlations developed through the different values of offsets. The above analysis would not be possible using standard traffic engineering techniques (e.g. the Highway Capacity Manual).

There is still a weak understanding on how one can characterize the breakdown dynamics and congestion spreading phenomena of traffic flows in these types of urban networks. While cascading phenomena are present in many types of physical or social systems (financing, interactions) city traffic has interesting irregularities that should be studied. We should identify the relative traffic variables that would allow a better prediction of the severity of developing and spreading of traffic congestion. This will provide clearer insights about the large variations of traffic congestion from one day to another even if demand profiles are similar.

These results can be of great importance to practitioners and city managers to unveil simple and robust signal timing planning in such a way that maximizes the network capacity and/or the density range of the capacity. The results of this paper can be utilized to develop efficient hierarchical control strategies for heterogeneously congested cities. A network can be partitioned in homogeneous regions (with small spatial variance of congestion distribution) and optimal control methodologies can identify the inter-transfers between regions of a city to maximize the system output, as expressed by number of trip endings (see for example Ji and Geroliminis (in press) for partitioning and Daganzo (2007), Geroliminis

and Daganzo (2007) or Haddad and Geroliminis (in press) for optimal control). The main logic of the strategies is that they try to decrease the inflow in regions with points in the decreasing part of an MFD and high demand for trip destinations. Given the estimated values from this task, the analysis of the current paper can identify signal parameters in the individual regions of a city to move traffic smoothly at the desired flows, without concentrating a large number of vehicles at the boundaries of the regions. By restricting access to congested cities, a city manager can significantly improve system output, highlighting the importance of a reliable estimator of subnetwork/route capacity. While there are vast contributions in traffic control problems for freeways through ramp metering, the area of control for large urban regions or mixed networks still remains a challenge. Our research provides tools to shed some light towards this direction.

Current extensions of this paper are investigating the network capacity and MFD patterns for cities with more complex structure (multiple modes of traffic competing for the same urban space). A difficult problem to address is how the redistribution of urban space between cars and more efficient modes can improve passenger network flows.

References

Buisson, C., Ladier, C., 2009. Exploring the impact of homogeneity of traffic measurements on the existence of macroscopic fundamental diagrams. Transportation Research Record 2124, 127–136.

Courbon, T., Leclercq, L., 2011. Cross-comparison of macroscopic fundamental diagram estimation methods. Procedia – Social and Behavioral Sciences 20 (0), 417–426 (the State of the Art in the European Quantitative Oriented Transportation and Logistics Research 14th Euro Working Group on Transportation & 26th Mini Euro Conference & 1st European Scientific Conference on Air Transport).

Daganzo, C., 2007. Urban gridlock: macroscopic modeling and mitigation approaches. Transportation Research Part B 41 (1), 49–62.

Daganzo, C., Menendez, M., 2005. A variational formulation of kinematic waves: bottleneck properties and examples. In: Proceedings of the 16th International Symposium on Transportation and Traffic Theory (ISTTT). pp. 345–364.

Daganzo, C.F., 2005. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transportation Research Part B 39 (2), 187–196.

Daganzo, C.F., Gayah, V.V., Gonzales, E.J., 2011. Macroscopic relations of urban traffic variables: bifurcations, multivaluedness and instability. Transportation Research Part B 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 42 (9), 771–781.

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

Geroliminis, N., Daganzo, C., 2007. Macroscopic modeling of traffic in cities. In: The 86th Transportation Research Board Annual Meeting. No. 07-0413. Washington, DC.

Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transportation Research Part B 42 (9), 759–770.

Geroliminis, N., Sun, J., 2011a. Hysteresis phenomena of a macroscopic fundamental diagram in freeway networks. Transportation Research Part A 45 (9), 966–979 (Paper published in the 19th International Symposium on Transportation and Traffic Theory – ISTTT).

Geroliminis, N., Sun, J., 2011b. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transportation Research Part B 45 (3), 605–617.

Godfrey, J.W., 1969. The mechanism of a road network. Traffic Engineering and Control 11 (7), 323–327.

Haddad, J., Geroliminis, N., in press. On the stability of traffic perimeter control in two-region urban cities. Transportation Research Part B, http://dx.doi.org/10.1016/j.trb.2012.04.004.

Helbing, D., 2009. Derivation of a fundamental diagram for urban traffic flow. The European Physical Journal B 70 (2), 229–241.

Herman, R., Prigogine, I., 1979. A two-fluid approach to town traffic. Science 204 (4389), 148–151.

Ji, Y., Daamen, W., Hoogendoorn, S., Hoogendoorn-Lanser, S., Qian, X., 2010. Investigating the shape of the macroscopic fundamental diagram using simulation data. Transportation Research Record 2161, 40–48.

Ji, Y., Geroliminis, N., in press. On the spatial partitioning of urban transportation networks. Transportation Research Part B, http://dx.doi.org/10.1016/j.trb.2012.08.005.

Laval, J.A., Leclercq, L., 2008. Microscopic modeling of the relaxation phenomenon using a macroscopic lane-changing model. Transportation Research Part B 42 (6), 511–522.

Leclercq, L., Laval, J.A., Chevallier, E., 2007. The lagrangian coordinates and what it means for first order traffic flow models. In: Proceedings of the 17 International Symposium on Transportation and Traffic Theory, pp. 735–753.

Mazloumian, A., Geroliminis, N., Helbing, D., 2010. The spatial variability of vehicle densities as determinant of urban network capacity. Philosophical Transactions of the Royal Society A 368 (1928), 4627–4647.

Saberi, M., Mahmassani, H., 2012. Exploring properties of network-wide flow-density relations in a freeway network. In: The 91st Transportation Research Board Annual Meeting. No. 12-2203. Washington DC.

Wardrop, J.G., 1952. Some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineers, Part II 1 (2), 325–378.