

Chapter 3

First-Order Macroscopic Traffic Models



3.1 Macroscopic Modelling Aspects

All macroscopic models, both of first-order type and of higher orders, describe the evolution of aggregate quantities referred to the traffic system over time. This means that two independent variables are involved, i.e. space and time. In *continuous traffic models*, these independent variables are assumed to be continuous, while they are discretised in *discrete traffic models*. In this latter case, a freeway stretch is divided into a number of small road portions, and the time horizon is subdivided into a given number of time intervals.

Let us now introduce the proper notation of macroscopic traffic models, specifically differentiated for the continuous and the discrete case.

3.1.1 The Continuous Case

Referring to a generic location x (in a given road, possibly composed of several lanes) and time t , the main aggregate variables considered in continuous macroscopic traffic models are:

- $\rho(x, t)$ is the traffic density [veh/km];
- $v(x, t)$ is the average speed [km/h];
- $q(x, t)$ is the traffic flow [veh/h].

A first relation constituting the basis of every macroscopic model is the *hydrodynamic equation*, which computes the flow as the product of mean speed and density, i.e.

$$q(x, t) = \rho(x, t)v(x, t) \tag{3.1}$$

A second relation is the *continuity equation* or *conservation equation*, directly derived from the conservation law of vehicle flows and expressed as

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial q(x, t)}{\partial x} = 0 \quad (3.2)$$

All the continuous macroscopic traffic models are based on (3.1) and (3.2) and differ for the other equations which relate the variables $\rho(x, t)$, $v(x, t)$ and $q(x, t)$. This chapter and the following one will introduce the most important continuous macroscopic models, respectively, of first-order and second-order type (see in particular Sects. 3.2 and 4.1). The interested reader can find an overview on continuous traffic models in [1].

As already discussed in Sect. 2.1.3, the theoretical relation between density and flow in steady-state conditions is the so-called *Fundamental Diagram*. This is a relation $Q(\rho(x, t))$, which has to satisfy the following conditions

$$Q(0) = 0, \quad Q(\rho^{\max}) = 0, \quad \left. \frac{dQ(\rho)}{d\rho} \right|_{\rho=\rho^{\text{cr}}} = 0 \quad (3.3)$$

where ρ^{cr} is the *critical density* [veh/km] and ρ^{\max} is the *jam density* [veh/km]. Moreover, q^{\max} is the *capacity* [veh/h].

Analogously, the steady-state relation between mean traffic speed and density is denoted with $V(\rho(x, t))$ and must satisfy the following conditions

$$V(0) = v^f, \quad V(\rho^{\max}) = 0, \quad \frac{dV(\rho)}{d\rho} \leq 0 \quad (3.4)$$

where v^f indicates the *free-flow speed* [km/h].

Different shapes of these steady-state relations have been proposed in the literature. The first types of diagrams were introduced by Greenshields in 1935 [2] and correspond to a linear form for $V(\rho(x, t))$ and a parabolic form for $Q(\rho(x, t))$, i.e.

$$\begin{aligned} V(\rho(x, t)) &= v^f \left[1 - \frac{\rho(x, t)}{\rho^{\max}} \right] \\ Q(\rho(x, t)) &= \rho(x, t) v^f \left[1 - \frac{\rho(x, t)}{\rho^{\max}} \right] \end{aligned} \quad (3.5)$$

In case relations (3.5) are applied, it holds by definition that $\rho^{\text{cr}} = \frac{1}{2}\rho^{\max}$ and $q^{\max} = \frac{1}{4}v^f\rho^{\max}$. An example of steady-state relations of type (3.5) is shown in Fig. 3.1.

Other possible shapes, widely used in the literature and adopted in the next chapters of this book, are

$$\begin{aligned} V(\rho(x, t)) &= v^f \exp \left[-\frac{1}{a} \left(\frac{\rho(x, t)}{\rho^{\text{cr}}} \right)^a \right] \\ Q(\rho(x, t)) &= \rho(x, t) v^f \exp \left[-\frac{1}{a} \left(\frac{\rho(x, t)}{\rho^{\text{cr}}} \right)^a \right] \end{aligned} \quad (3.6)$$

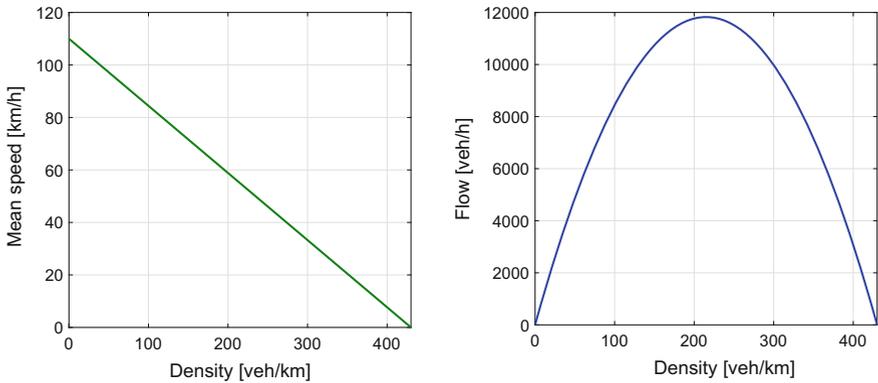


Fig. 3.1 Example of steady-state relations of type (3.5) with $v^f = 110$ [km/h], $\rho^{\max} = 430$ [veh/km]

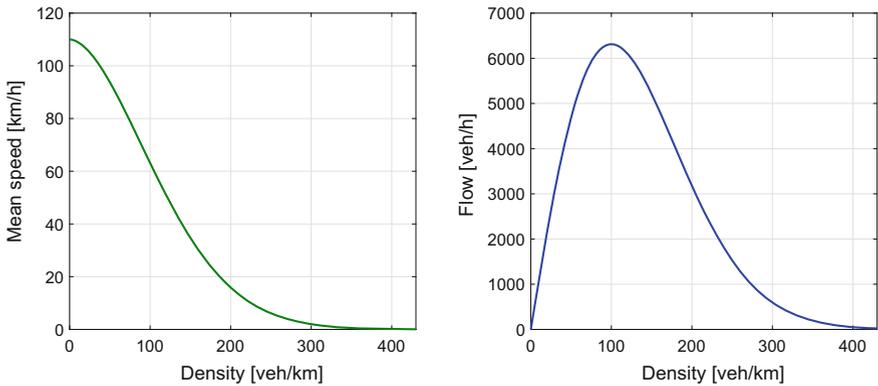


Fig. 3.2 Example of steady-state relations of type (3.6) with $v^f = 110$ [km/h], $\rho^{\text{cr}} = 100$ [veh/km], $a = 1.8$

where $a > 0$ is a suitable parameter. Note that these exponential relations do not meet the conditions $V(\rho^{\max}) = 0$ and $Q(\rho^{\max}) = 0$, respectively, in (3.4) and (3.3), but it holds that the values of $V(\rho^{\max})$ and $Q(\rho^{\max})$ in (3.6) are very small, hence in some way approximating conditions $V(\rho^{\max}) = 0$ and $Q(\rho^{\max}) = 0$. An example of steady-state relations of type (3.6) is reported in Fig. 3.2.

Finally, other common shapes of steady-state relations are

$$\begin{aligned}
 V(\rho(x, t)) &= v^f \left[1 - \left(\frac{\rho(x, t)}{\rho^{\max}} \right)^l \right]^m \\
 Q(\rho(x, t)) &= \rho(x, t) v^f \left[1 - \left(\frac{\rho(x, t)}{\rho^{\max}} \right)^l \right]^m
 \end{aligned} \tag{3.7}$$

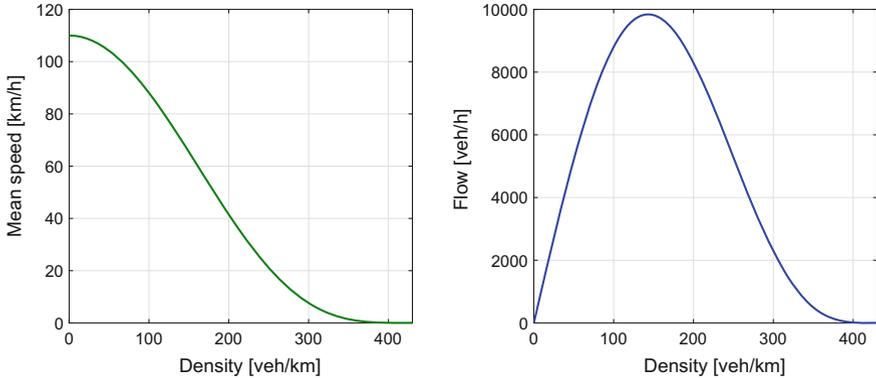


Fig. 3.3 Example of steady-state relations of type (3.7) with $v^f = 110$ [km/h], $\rho^{\max} = 430$ [veh/km], $l = 2$, $m = 4$

where $l > 0$, $m > l$ are parameters. Note that (3.7) are very general and can represent most of the shapes reported in the literature, such as (3.6), for given values of l and m [3]. Figure 3.3 provides an example of steady-state relations of type (3.7).

3.1.2 The Discrete Case

In case of discrete macroscopic traffic models, space is divided into N portions of length L [km] and time is discretised into K time intervals of duration T [h]. Let us denote with $i = 1, \dots, N$ the generic road portion (in some models called *cell* and in others called *section*), and with $k = 0, \dots, K$ the generic time step. In some models, the space discretisation is not uniform, hence each portion i may have a different length L_i , $i = 1, \dots, N$.

Referring to a generic portion i (in a road which can be composed of several lanes) and time step k , the main aggregate variables to be considered are:

- $\rho_i(k)$ is the traffic density at time kT [veh/km];
- $v_i(k)$ is the mean speed at time kT [km/h];
- $q_i(k)$ is the traffic flow during time interval $[kT, (k+1)T)$ [veh/h].

The hydrodynamic and continuity equations, in the discrete case, become

$$q_i(k) = \rho_i(k)v_i(k) \quad (3.8)$$

$$\rho_i(k+1) = \rho_i(k) + \frac{T}{L} [I_i(k) - O_i(k)] \quad (3.9)$$

where $I_i(k)$ is the traffic flow entering portion i during time interval $[kT, (k+1)T)$ [veh/h], and $O_i(k)$ is the traffic flow exiting portion i in the same time interval [veh/h].

Analogously to the continuous case, steady-state relations among flow, density and mean speed can be defined. In particular, relation (3.5) becomes

$$\begin{aligned} V(\rho_i(k)) &= v^f \left[1 - \frac{\rho_i(k)}{\rho^{\max}} \right] \\ Q(\rho_i(k)) &= \rho_i(k) v^f \left[1 - \frac{\rho_i(k)}{\rho^{\max}} \right] \end{aligned} \quad (3.10)$$

Similarly, relation (3.6) can be written as

$$\begin{aligned} V(\rho_i(k)) &= v^f \exp \left[-\frac{1}{a} \left(\frac{\rho_i(k)}{\rho^{\text{cr}}} \right)^a \right] \\ Q(\rho_i(k)) &= \rho_i(k) v^f \exp \left[-\frac{1}{a} \left(\frac{\rho_i(k)}{\rho^{\text{cr}}} \right)^a \right] \end{aligned} \quad (3.11)$$

and (3.7) as

$$\begin{aligned} V(\rho_i(k)) &= v^f \left[1 - \left(\frac{\rho_i(k)}{\rho^{\max}} \right)^l \right]^m \\ Q(\rho_i(k)) &= \rho_i(k) v^f \left[1 - \left(\frac{\rho_i(k)}{\rho^{\max}} \right)^l \right]^m \end{aligned} \quad (3.12)$$

In some models, a different steady-state relation is considered for each road portion. In these cases, the parameters of the previous relations can be indexed with i , namely, v_i^f , ρ_i^{\max} , ρ_i^{cr} , a_i , l_i , m_i , $i = 1, \dots, N$.

3.2 Continuous First-Order Models

The first macroscopic traffic model was developed by Lighthill and Whitham [4] and by Richards [5] in the 50s and is now known as the *Lighthill–Whitham–Richards* (LWR) model. The basic assumption of the LWR model is that vehicles adjust their speeds instantaneously to the value given by the steady-state relation depending on the present density. This model has been extended to consider boundary conditions, sources and inhomogeneities, as well as to represent traffic networks, as it will be described in the following subsections.

Most of the results on the LWR model have been obtained considering that it belongs to the class of conservation laws, for which a thorough theory has been developed by mathematicians (see, e.g. the books [6–10]). Considering more specifically the LWR model, especially its application for traffic networks, the interested

reader can find more details in [11, 12], where all the related mathematical aspects are discussed in detail.

The LWR model presents several limitations. For instance, it does not contain any inertial effects, since it assumes that vehicles adjust their speeds instantaneously. This can produce unrealistically high accelerations or decelerations of vehicles. Moreover, it systematically predicts that the output flow from a congested area is equal to the capacity flow, if the portion of road downstream is not congested. This is in contrast with the capacity drop phenomenon observed in real-world traffic networks, as discussed in Sect. 2.2.4. Other qualitative considerations on first-order macroscopic models are included in [13].

3.2.1 The LWR Model

The LWR model is based on the assumption that the traffic flow instantaneously follows the density according to the Fundamental Diagram. The LWR model is then given by (3.1), (3.2), and the following relation:

$$v(x, t) = V(\rho(x, t)) \quad (3.13)$$

or, alternatively,

$$q(x, t) = Q(\rho(x, t)) \quad (3.14)$$

Hence, the LWR model can be rewritten as

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial(Q(\rho(x, t)))}{\partial x} = 0 \quad (3.15)$$

or

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial(\rho(x, t)V(\rho(x, t)))}{\partial x} = 0 \quad (3.16)$$

The LWR model belongs to the class of *first-order models*, in the sense that it captures the dynamics of a single variable, namely, the traffic density. Moreover, this model, in its original version, makes some assumptions on the shape of the Fundamental Diagram. Specifically, it assumes that $Q(\rho(x, t))$ is a C^2 function, is strictly concave, and ensures that $Q(0) = 0$ and $Q(\rho^{\max}) = 0$, as in (3.3). According to these assumptions, (3.15) belongs to the class of *hyperbolic conservation laws*. The theory of systems of conservation laws has been extensively studied in the literature with particular attention to the problem of well-posedness, as done, for instance, in [14–18].

A very peculiar aspect associated with conservation laws is the generation of *discontinuities*. Referring in particular to the LWR model in the traffic case, the discontinuities resulting from the solution of the LWR model satisfactorily reproduce

the shock waves which can be actually observed in traffic systems. These aspects were discussed in detail in [4, 5], where the wave theory is applied and the propagation of kinematic waves is discussed in detail, with reference to the real behaviour of road traffic systems. In [4], some preliminary comments to the traffic behaviour at road junctions are reported as well.

From a mathematical point of view, a relevant effort was put by researchers on solving the so-called LWR *initial value problem*, given by the conservation law (3.15) with a specified initial condition for the density $\rho(x, 0) = \rho_I(x)$. Referring to the general theory on conservation laws, it can be easily shown that the solution to this initial value problem can produce discontinuities in finite time, even in case of continuous initial conditions. This can be shown by applying the method of *characteristics*, allowing to rewrite the partial differential equation of the LWR model as a system of ordinary differential equations; the characteristics can be seen as lines in the (x, t) plane, starting from space–time points where initial conditions are known, along which the solution remains constant. If these lines do not intersect, the solution is unique; if instead they intersect, this means that there is a discontinuity (a shock) in the solution, and this is what normally happens with the LWR model. In this latter case, weak solutions must be dealt with (see, for instance, [11] for further mathematical details).

One of the most interesting cases to be analysed, especially when referring to traffic applications, is the solution of the LWR initial value problem in case the initial condition for the density $\rho_I(x)$ is piecewise constant. This corresponds, for instance, to the presence of vehicles waiting in front of a red traffic light: the density after the traffic light is low, while the density before is high. The opposite example is the case of queue formation for a red traffic light or for an accident in a freeway stretch: there is a point in space after which the density is very high and before which free-flow traffic conditions are present. The initial value problem in case of a discontinuous initial condition is called *Riemann problem*. Let us consider specifically the Riemann problem for (3.15) with the initial condition expressed as

$$\rho(x, 0) = \rho_I(x) = \begin{cases} \rho^- & \text{if } x < 0 \\ \rho^+ & \text{if } x > 0 \end{cases} \quad (3.17)$$

By applying the general results on conservation laws, it can be shown that this Riemann problem has not unique solution. The conventional mathematical approach to solve this problem is devoted to look for *entropy-admissible solutions* (see, for instance, [11] for a rigorous definition of this type of solutions), which present good properties, such as the uniqueness and the fact that they depend continuously on initial data. In [19], an interesting discussion about the choice of adopting entropy solutions for the LWR model is reported: the author explains, through an example, that the choice of the entropy solution is a mathematical sound choice, which guarantees existence, uniqueness and continuous dependency on initial conditions, but in some cases, these entropy solutions are not the best choice in order to provide a realistic behaviour of traffic.

The entropy-admissible solution for the Riemann problem for (3.15) with initial condition (3.17) can be written by distinguishing two cases:

1. if $\rho^- < \rho^+$, i.e. $\frac{dQ(\rho^-)}{d\rho} > \frac{dQ(\rho^+)}{d\rho}$ for the considered assumptions on $Q(\rho(x, t))$, the entropy-admissible solution is given by the *shock wave* expressed as

$$\rho(x, t) = \begin{cases} \rho^- & \text{if } x < \lambda t \\ \rho^+ & \text{if } x > \lambda t \end{cases} \quad (3.18)$$

where λ is obtained by applying the so-called Rankine–Hugoniot condition and is given by

$$\lambda = \frac{Q(\rho^+) - Q(\rho^-)}{\rho^+ - \rho^-} \quad (3.19)$$

This solution corresponds to a discontinuity, in which the density abruptly changes from ρ^- to ρ^+ , propagating in space and time with speed λ , which then represents the *shock front propagation speed*;

2. if $\rho^- > \rho^+$, i.e. $\frac{dQ(\rho^-)}{d\rho} < \frac{dQ(\rho^+)}{d\rho}$, the entropy-admissible solution is given by the *rarefaction wave* expressed as

$$\rho(x, t) = \begin{cases} \rho^- & \text{if } x < \frac{dQ(\rho^-)}{d\rho} t \\ \left[\frac{dQ(\frac{x}{t})}{d\rho} \right]^{-1} & \text{if } \frac{dQ(\rho^-)}{d\rho} t < x < \frac{dQ(\rho^+)}{d\rho} t \\ \rho^+ & \text{if } x > \frac{dQ(\rho^+)}{d\rho} t \end{cases} \quad (3.20)$$

In this case, the solution is continuous, i.e. the propagation of the density in space and time occurs in a smooth way.

It is worth noting that, in case 1, the characteristics on the (x, t) plane overlap, as shown in the left graph of Fig. 3.4. Hence, the solution implies a discontinuity, which is highlighted in the right graph of Fig. 3.4, where the shock wave is represented by a dashed green line. Figure 3.5 shows instead the characteristics in case 2: they do not overlap; hence, there is a region in the (x, t) plane which appears to be empty. In that region, called *expansion fan*, characteristics are rays of constant density originating at $x = 0$ in order to guarantee continuity of the solution (see the right graph of Fig. 3.5).

The solution described so far holds for the case of strictly concave Fundamental Diagram. Nevertheless, also the case in which the Fundamental Diagram is *non-concave* can be interesting for real traffic applications. This case was treated in some research papers, such as in [20, 21].

Another version of the LWR model was introduced to overcome the fact that the LWR model produces discontinuities in finite time, leading to the so-called *LWR model with viscosity*. In this model, a viscosity term is added to (3.15), i.e.

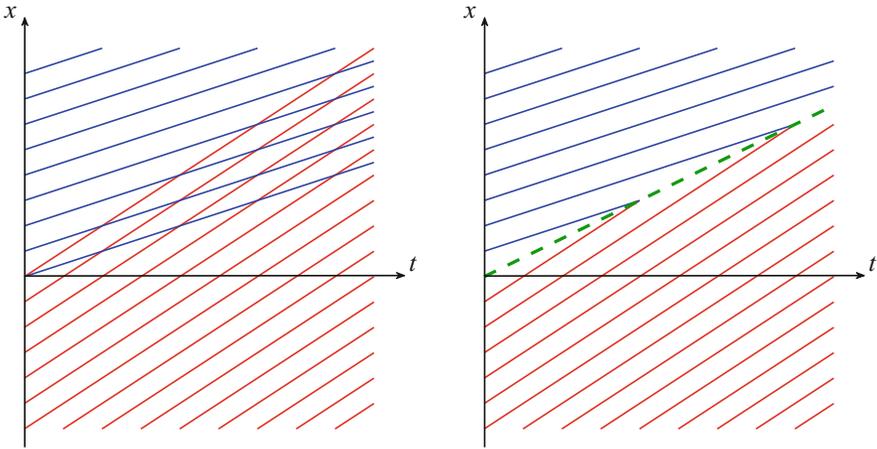


Fig. 3.4 Characteristics in case 1: shock wave

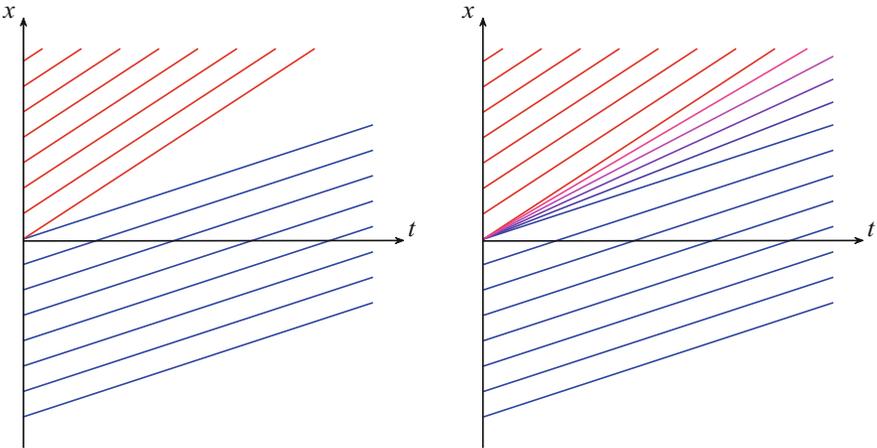


Fig. 3.5 Characteristics in case 2: rarefaction wave

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial(Q(\rho(x, t)))}{\partial x} = \mu \frac{\partial^2 \rho(x, t)}{\partial x^2} \tag{3.21}$$

and the discontinuities in the solution are eliminated. Nevertheless, in [11], it is shown that this model is not realistic to describe the traffic flow evolution.

3.2.2 *The LWR Model with Boundary Conditions, Sources and Inhomogeneities*

In the initial value problem described in Sect. 3.2.1, no boundary conditions are defined for the boundaries of the space domain. Clearly, this is somewhat unrealistic for a freeway traffic system, since normally a given road stretch is considered and the traffic conditions at the beginning and at the end of the stretch must be taken into account. To consider boundary conditions, an *initial-boundary value problem* must be addressed, in which the conservation law must satisfy, not only an initial condition but also the boundary conditions. Also, the initial-boundary value problem for the general class of hyperbolic conservation laws has been widely studied by mathematicians, with specific attention to the well-posedness of the problem, developing conditions for the existence and unicity of the solution (see, e.g. [18, 22, 23]).

In the specific case of the LWR model for traffic systems, the initial-boundary value problem is given by (3.15), with the initial condition $\rho(x, 0) = \rho_I(x)$ and boundary conditions that can be expressed in different ways. The boundary conditions can be related to the values of the density at the boundaries, i.e.

$$\rho(0, t) = \rho_0(t), \quad \rho(x_L, t) = \rho_L(t) \quad (3.22)$$

where $x = 0$ and $x = x_L$ indicate the initial and final location of the considered freeway stretch. Another possibility is that the boundary conditions are related to the values of the flow at the boundaries. This latter case is more realistic for many traffic systems, since traffic sensors generally provide measurements of traffic flows, whereas it is more difficult to estimate the values of the density in specific locations. In this case, the boundary conditions are given by

$$q(\rho(0, t)) = q_0(t), \quad q(\rho(x_L, t)) = q_L(t) \quad (3.23)$$

Some works in the literature deal with the initial-boundary value problem for the LWR model specifically referred to the case of freeway traffic. For instance, in [24], the boundary conditions are given in terms of traffic density, and the existence and uniqueness of a weak solution are proved. Also, the proposed numerical scheme is applied to a freeway scenario with data of the Interstate-80 Eastbound in West Berkeley and Emeryville, U.S. Another work dealing with the initial-boundary value problem for the LWR model is [25], referred to freeway stretches. In [25], the boundary conditions refer to the time-dependent flow entering a specific location, namely, $x = 0$. Moreover, some constraints on the flow at another specific location are included, namely, at $x = x_C$, modelling the presence of toll gates, construction sites or the occurrence of accidents, which limit the traffic flow. The boundary conditions are then expressed as

$$q(\rho(0, t)) = q_0(t), \quad q(\rho(x_C, t)) \leq \bar{q}_C(t) \quad (3.24)$$

where $\bar{q}_C(t)$ is the maximum flow allowed at $x = x_C$. The well-posedness result provided in [25] allows also to prove the existence of optimal management strategies for freeway traffic systems.

The LWR model described so far does not take into account the presence of on-ramps and off-ramps, which are instead a very important issue in modelling freeway stretches. In order to consider entrances of vehicles from on-ramps, exits from off-ramps, as well as local changes of the traffic flow due to inhomogeneities of the road, the LWR model must be written as a *conservation law with source* or *inhomogeneous conservation law*, i.e.

$$\frac{\partial \rho(x, t)}{\partial t} + \frac{\partial(Q(\rho(x, t)))}{\partial x} = s(x, t, \rho) \quad (3.25)$$

where $s(x, t, \rho)$ is the source term. Well-posedness results and numerical investigations for the inhomogeneous LWR model are presented in [26], where also second-order models with source terms are analysed.

Another interesting aspect to be included in the LWR model, relevant especially for real contexts, is related to consider the case in which the Fundamental Diagram depends explicitly and (sometimes discontinuously) on x and on t . This can allow to model intersections, sections with variable number of lanes, portions of the road with local and temporary variations of the parameters (e.g. capacity or free-flow speed). If the Fundamental Diagram only varies depending on time, the corresponding LWR model is said to present *time inhomogeneity*, whereas it has *space inhomogeneity* if the Fundamental Diagram only depends on space.

The LWR model with space inhomogeneity has been studied deeply and the well-posedness of the associated initial value problem (both for continuous and discontinuous dependence of the Fundamental Diagram on x) has been proven [19, 27]. The case of space–time inhomogeneity has been studied more recently, for example, in [28].

An alternative option for the solution of the LWR is given by the *Hamilton–Jacobi theory*, adopted, for instance, in [29, 30]. According to this theory, a *Lagrangian* approach, which is trajectory-based, is adopted, in contrast with the standard *Eulerian* framework used to solve conservation laws. This approach can assume particular relevance especially for the new type of sensors that are more and more widespread in freeway networks, i.e. mobile sensors which travel inside the domain along trajectories, providing *internal conditions* for the problem, in addition with boundary conditions provided by standard traffic sensors.

3.2.3 The LWR Model on Networks

In order to represent large-scale freeway systems, the LWR model has been extended to the case of *networks*, in which each road is modelled with the LWR model and specific conditions must be defined for the junctions where roads intersect. The first

work in this direction was reported in [31], where a network of unidirectional roads is seen as a connected directed *graph*, with edges modelling the roads and vertices corresponding to junctions. The *junctions* play a key role in these network models, since at junctions the system is underdetermined even if the conservation of cars is taken into account; in other words, in order to obtain a well-defined solution, it is necessary to specify the distribution of vehicles at the junctions. In [31], the Riemann problem for the considered system is solved by *maximising the flow* at each intersection, and the existence of a solution to the general Cauchy problem is proven.

In [32], the road network is modelled as a graph, similarly to the case proposed in [31], but different conditions at junctions are taken into account. Specifically, it is assumed that there are some prescribed *preferences* of drivers, i.e. the traffic from incoming roads is distributed on outgoing roads according to fixed coefficients, and, by following these preferences, the drivers behave in order to maximise the traffic flows. Considering this model, the authors of [32] prove the existence of solutions to the Cauchy problem and show that the Lipschitz continuous dependence by initial data holds only under specific assumptions. Some other research works have dealt with developments of the LWR model on networks (see, e.g. [33, 34]), also considering specific types of junctions. For instance, the authors of [35] analyse the case of a T-junction, in which the interactions among incoming and outgoing flows are explicitly modelled. In [36], the specific case of freeways is addressed, and the considered junction is composed of the mainstream, an off-ramp and an on-ramp, this latter being modelled as a buffer of infinite capacity.

Some studies have also considered the case of *nodes with buffer*, i.e. the case in which there is a dynamics inside the junction, generally described by ordinary differential equations depending on incoming and outgoing flows. For instance, in [37], the storage capacity of the junctions is taken into account by using a reformulation of intersection models in terms of supply and demand functions. Similarly to [37], in [38], the solution of the Riemann problem at the node is provided and existence and well-posedness of solutions to the Cauchy problem are proven. A multi-buffer model is studied in [39], where a set of buffers, one for each outgoing road, is considered, allowing to correctly respect the preferences of drivers.

3.3 Discrete First-Order Models

Different numerical methods for non-linear conservation laws have been studied by researchers since many decades. While approximating a partial differential equation with a finite-difference equation, it is of course interesting to evaluate the error due to this approximation and to study relevant properties such as convergence and stability of the numerical method (see, e.g. [7, 40] for a detailed description of numerical methods for conservation laws). Moreover, to discretise partial differential equations, one can use both explicit and implicit numerical methods. With the *explicit* solution scheme, it is possible to explicitly express the dependence of each variable in the

current time step on the variables in previous time steps. In the *implicit* case, instead, it is necessary to solve a system of equations involving both current and past values of the different variables (see, e.g. [41]). Implicit formulas are typically more stable than explicit ones, but harder to implement [42].

Referring to the specific case of traffic, i.e. to the LWR model, different finite-difference approximations have been proposed in the last decades. According to these numerical methods, the road space is divided into portions of finite length, time is discretised into time intervals of equal duration, and the partial differential equation of the LWR model is transformed into a finite-difference equation.

The most famous discretised version of the LWR model is the *Cell Transmission Model* (CTM), presented for the first time by Daganzo in [43, 44], making reference to a one-way road without any intermediate entrances or exits, and then extended in [45] for traffic networks with three-legged junctions, hence allowing to model on-ramps, off-ramps, freeway intersections and so on. According to the CTM, the discrete portions of the road are called *cells*, and two quantities are associated with the intersection between two cells, i.e. a *sending* function depending on the density before the intersection, and a *receiving* function depending on the density downstream the intersection.

In [43, 44], it is shown that the CTM is a discrete equivalent of the classical LWR model, both in case of continuous density and in presence of discontinuities. Moreover, the author of [43, 44] argues that the CTM could capture real-life features, such as stop-and-go phenomena, that the LWR theory is not able to model. This analysis is carried out by considering the specific case of triangular or trapezoidal Fundamental Diagram $Q(\rho)$, but it is asserted that the considerations reported in those papers can be generalised to other shapes of $Q(\rho)$. In [46], the propagation of disturbances of the CTM is also analysed and an asymptotic formula for the errors introduced by the finite difference approximation is presented.

A very interesting analysis of discretisation of first-order traffic flow models was conducted by Lebacque in [19], where he focuses on a specific numerical method, that is the so-called *Godunov scheme* [47]. This is a conservative finite-volume method which solves Riemann problems at each cell interface forward in time. In [19], it is shown that the CTM corresponds to the application of the Godunov scheme to the LWR model. In particular, the sending and receiving functions computed at the intersection between subsequent cells in the CTM are equivalent to the values of the flow at the singularity in the solution of the Riemann problem in the Godunov scheme. In [19], Lebacque introduces the terminology *demand* and *supply*, respectively, for the sending and receiving functions; this terminology is presently the most widespread when using the CTM and it is also the one adopted in this book.

By applying the Godunov scheme, a condition for the space discretisation L and the time discretisation T is also derived, which can be expressed as

$$T \max_{\rho \in [0, \rho^{\max}]} \left| \frac{dQ(\rho)}{d\rho} \right| \leq L \quad (3.26)$$

In [19], different shapes for the Fundamental Diagram $Q(\rho)$ (and consequently for the demand and supply functions) are investigated, also considering the case in which $Q(\rho)$ includes discontinuities, and it is argued that the concepts of supply and demand can provide an effective tool also for modelling intersections and networks.

Among the different discretisations of the LWR model proposed in the literature, it is worth mentioning also the *Link Transmission Model* (LTM), introduced for the first time in [48]. In the LTM, the evolution of traffic on a generic road network is represented in terms of the cumulative number of vehicles that pass the initial and final locations of each link at each time step. Hence, the numerical procedure characterising the LTM only requires calculations at the link boundaries, as in [49], instead of at each cell boundary, as in the CTM. This results in a computational advantage compared with the CTM. More efficient numerical schemes have been developed starting from the LTM, such as the iterative algorithm described in [50].

In the following subsections, we will focus on the CTM and its extensions, both for a freeway stretch and for a freeway network, since this is surely the most widespread first-order model in the traffic control engineering community, and, hence, of particular interest for the purposes of the present book.

3.3.1 The CTM for a Freeway Stretch

Let us start from the CTM for a freeway stretch including on-ramps and off-ramps. Note that the CTM described hereafter is derived from the original version proposed by Daganzo in [43–45], but it is presented with a different mathematical notation and nomenclature in order to conform to the notation and model classification adopted in this book.

As previously introduced in the general notation of a discrete traffic model, let N be the number of cells and K the number of time intervals. Let T denote the sample time [h] and L the length of each cell [km]. Moreover, in the CTM, on-ramps and off-ramps are assumed to be present at the interface between two subsequent cells.

For each cell $i = 1, \dots, N$, and for each time step $k = 0, \dots, K$, let us define the following quantities:

- $\rho_i(k)$ is the traffic density of cell i at time kT [veh/km];
- $\Phi_i^+(k)$ is the total flow entering cell i during time interval $[kT, (k+1)T)$ [veh/h];
- $\Phi_i^-(k)$ is the total flow exiting cell i during time interval $[kT, (k+1)T)$ [veh/h];
- $\phi_i(k)$ is the mainstream (interface) flow entering cell i from cell $i-1$ during time interval $[kT, (k+1)T)$ [veh/h];
- $r_i(k)$ is the flow entering cell i from the on-ramp during time interval $[kT, (k+1)T)$ [veh/h];
- $s_i(k)$ is the flow exiting cell i through the off-ramp during time interval $[kT, (k+1)T)$ [veh/h];
- $\beta_i(k) \in [0, 1)$ is the split ratio of cell i during time interval $[kT, (k+1)T)$;

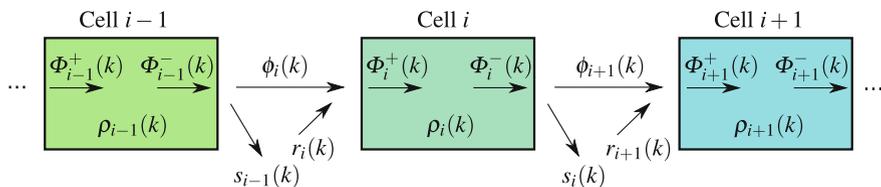


Fig. 3.6 Sketch of the division of the freeway stretch into cells and the relative notation in the CTM

- $D_i(k)$ is the demand of cell i (i.e. flow that can be sent from cell i to cell $i + 1$) during time interval $[kT, (k + 1)T)$ [veh/h];
- $S_i(k)$ is the supply of cell i (i.e. flow that can be received by cell i from cell $i - 1$) during time interval $[kT, (k + 1)T)$ [veh/h];
- $D_i^{\text{ramp}}(k)$ is the demand of the on-ramp of cell i (i.e. flow that can be sent from the on-ramp into cell i) during time interval $[kT, (k + 1)T)$ [veh/h].

Figure 3.6 depicts a sketch of the subdivision of the freeway stretch into cells, with the main variables of the CTM.

The parameters of the CTM are as follows: v_i is the free-flow speed of cell i [km/h], w_i is the congestion wave speed of cell i [km/h], q_i^{max} is the capacity of cell i [veh/h], ρ_i^{max} is the jam density of cell i [veh/km], $p_i^{\text{ramp}} \in [0, 1]$ is the priority of the on-ramp flow with respect to the mainstream flow in cell i , $p_i \in [0, 1]$ is the priority of the mainstream flow with respect to the on-ramp flow in cell i , such that $p_i^{\text{ramp}} + p_i = 1$, $i = 1, \dots, N$.

The CTM is characterised by the following equations describing the traffic density, this latter being the state variable of dimension N :

$$\rho_i(k + 1) = \rho_i(k) + \frac{T}{L} [\Phi_i^+(k) - \Phi_i^-(k)] \quad (3.27)$$

where $i = 1, \dots, N$, $k = 0, \dots, K - 1$, and the total flows entering and exiting cell i are, respectively, given by

$$\Phi_i^+(k) = \phi_i(k) + r_i(k) \quad (3.28)$$

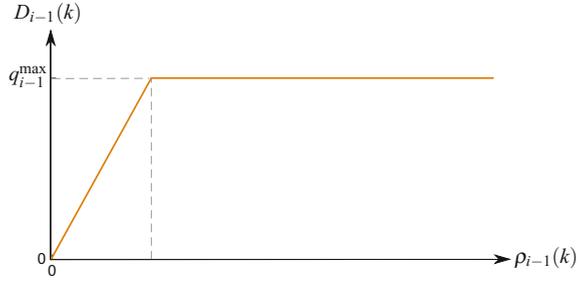
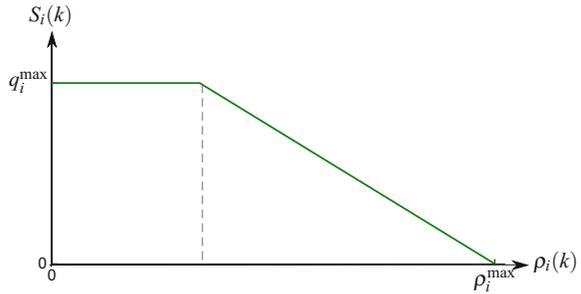
$$\Phi_i^-(k) = \phi_{i+1}(k) + s_i(k) \quad (3.29)$$

The flow exiting through the off-ramp is computed as

$$s_i(k) = \frac{\beta_i(k)}{1 - \beta_i(k)} \phi_{i+1}(k) \quad (3.30)$$

since $s_i(k) = \beta_i(k) \Phi_i^-(k) = \beta_i(k) [\phi_{i+1}(k) + s_i(k)]$.

Two important concepts of the CTM are the *demand* and the *supply*, associated with each cell. In particular, referring to the boundary between cell $i - 1$ and cell i ,

Fig. 3.7 Demand function in the CTM**Fig. 3.8** Supply function in the CTM

let us introduce the demand of cell $i - 1$, namely, $D_{i-1}(k)$, and the supply of cell i , namely, $S_i(k)$. The demand $D_{i-1}(k)$ is the flow that cell $i - 1$ could send to the next cell i during time interval $[kT, (k + 1)T)$, while the supply $S_i(k)$ is the flow that cell i could receive from cell $i - 1$ in the same time interval. These two quantities are computed as

$$D_{i-1}(k) = \min \left\{ (1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k), q_{i-1}^{\max} \right\} \quad (3.31)$$

$$S_i(k) = \min \left\{ w_i(\rho_i^{\max} - \rho_i(k)), q_i^{\max} \right\} \quad (3.32)$$

The demand and the supply are shown in Figs. 3.7 and 3.8, respectively, as functions of the density.

The *merge* between the on-ramp and the mainstream is analogous to the merge of two generic cells, as described in the original model proposed in [45]. In the generic case described in [45], a merge is given by two sending cells (characterised by two specific demands) and one receiving cell (characterised by a given supply); according to [45], the two sending cells send the maximum possible flow that the receiving cell is able to host. This merge model is adopted to compute the mainstream and on-ramp flows, since this latter case can be seen as a situation of two sending cells (the mainstream and the on-ramp) and one receiving cell downstream. In particular, for a given cell i during time interval $[kT, (k + 1)T)$, the demands of the sending cells are $D_{i-1}(k)$ and $D_i^{\text{ramp}}(k)$, while the supply of the receiving cell is $S_i(k)$.

Two cases must be distinguished, corresponding, respectively, to free-flow and congested conditions.

Free-Flow Case This is the case in which there is enough space for the two flows that want to enter cell i , i.e.

$$\begin{aligned} \text{If } D_{i-1}(k) + D_i^{\text{ramp}}(k) &\leq S_i(k) \\ \text{then } \phi_i(k) &= D_{i-1}(k), \quad r_i(k) = D_i^{\text{ramp}}(k) \end{aligned} \quad (3.33)$$

Congested Case The congested case is the opposite situation in which not all the flows that want to enter cell i can be received by it, i.e.

$$\begin{aligned} \text{If } D_{i-1}(k) + D_i^{\text{ramp}}(k) &> S_i(k) \\ \text{then } \phi_i(k) &= \text{mid} \{ D_{i-1}(k), S_i(k) - D_i^{\text{ramp}}(k), p_i S_i(k) \} \\ r_i(k) &= \text{mid} \{ D_i^{\text{ramp}}(k), S_i(k) - D_{i-1}(k), p_i^{\text{ramp}} S_i(k) \} \end{aligned} \quad (3.34)$$

where the function `mid` returns the middle value.

In order to better understand the merge model in the congested case, remind that the case $D_{i-1}(k) + D_i^{\text{ramp}}(k) \geq S_i(k)$ corresponds to a situation in which it is not possible to completely satisfy the demand $D_{i-1}(k)$ from the mainstream and the demand $D_i^{\text{ramp}}(k)$ from the on-ramp. Moreover, remind that parameters p_i and p_i^{ramp} model, respectively, the priority of the mainstream flow and the on-ramp flow in the merge and that $p_i^{\text{ramp}} + p_i = 1$.

The basic idea of the merge model is that the demand $D_{i-1}(k)$ has a ‘reserved’ flow equal to $p_i S_i(k)$, while the demand $D_i^{\text{ramp}}(k)$ has a ‘reserved’ flow of $p_i^{\text{ramp}} S_i(k)$. Another important assumption of the merge model proposed in [45] is that if one of the two demands is lower than the corresponding ‘reserved’ flow, the complementary flow will saturate the supply of the receiving cell.

By rewriting (3.34) as

$$\begin{aligned} \text{If } S_i(k) - D_i^{\text{ramp}}(k) &\leq p_i S_i(k) \leq D_{i-1}(k) \\ \text{then } \phi_i(k) &= p_i S_i(k), \quad r_i(k) = p_i^{\text{ramp}} S_i(k) \\ \text{If } p_i S_i(k) &\leq S_i(k) - D_i^{\text{ramp}}(k) \leq D_{i-1}(k) \\ \text{then } \phi_i(k) &= S_i(k) - D_i^{\text{ramp}}(k), \quad r_i(k) = D_i^{\text{ramp}}(k) \\ \text{If } S_i(k) - D_i^{\text{ramp}}(k) &\leq D_{i-1}(k) \leq p_i S_i(k) \\ \text{then } \phi_i(k) &= D_{i-1}(k), \quad r_i(k) = S_i(k) - D_{i-1}(k) \end{aligned} \quad (3.35)$$

it is possible to distinguish three sub-cases:

- if $D_{i-1}(k) \geq p_i S_i(k)$ and $D_i^{\text{ramp}}(k) \geq p_i^{\text{ramp}} S_i(k)$, then the ‘reserved’ flows are guaranteed, resulting in $\phi_i(k) = p_i S_i(k)$ and $r_i(k) = p_i^{\text{ramp}} S_i(k)$;
- if $D_{i-1}(k) \geq p_i S_i(k)$ and $D_i^{\text{ramp}}(k) \leq p_i^{\text{ramp}} S_i(k)$, i.e. the demand from the on-ramp is lower than the ‘reserved’ flow, all the demand from the on-ramp enters the cell

and the flow entering from cell $i - 1$ is obtained in order to saturate the supply $S_i(k)$, resulting in $\phi_i(k) = S_i(k) - D_i^{\text{ramp}}(k)$ and $r_i(k) = D_i^{\text{ramp}}(k)$;

- if $D_{i-1}(k) \leq p_i S_i(k)$ and $D_i^{\text{ramp}}(k) \geq p_i^{\text{ramp}} S_i(k)$, i.e. the mainstream demand is lower than the ‘reserved’ flow, all the mainstream demand enters the cell and the flow entering from the on-ramp is obtained in order to saturate the supply $S_i(k)$, resulting in $\phi_i(k) = D_{i-1}(k)$ and $r_i(k) = S_i(k) - D_{i-1}(k)$.

Note that, in any case, in congested situations the total flow entering cell i is given by $\Phi_i^+(k) = \phi_i(k) + r_i(k) = S_i(k)$.

Summarising, the CTM for a freeway with off-ramps and on-ramps in all the cells is given by (3.27)–(3.34). The *boundary conditions* are the demand in the cell before the first one, i.e. $D_0(k)$, the supply of the cell after the last one, i.e. $S_{N+1}(k)$, the on-ramp demands, i.e. $D_i^{\text{ramp}}(k)$, and the split ratios, i.e. $\beta_i(k)$, $i = 1, \dots, N$, $k = 0, \dots, K$.

Finally, let us consider the case in which some cells have *no off-ramps and no on-ramps*. To adapt the CTM previously described to the case in which some cells do not present any ramps, it is possible to fix $\beta_{i-1}(k) = 0$, $D_i^{\text{ramp}}(k) = 0$ and $p_i^{\text{ramp}} = 0$ in case there are not on-ramps and off-ramps between cell $i - 1$ and cell i , $i = 1, \dots, N$. In this way, it is assured that $r_i(k) = 0$ and $s_{i-1}(k) = 0$, $k = 0, \dots, K$. Note that in this case the interface flow can be computed as

$$\phi_i(k) = \min\{D_{i-1}(k), S_i(k)\} \quad (3.36)$$

according to the first CTM proposed in [43].

3.3.2 The CTM with On-Ramp Queue Dynamics

The CTM described in Sect. 3.3.1 considers a freeway stretch with on-ramps and off-ramps and models the dynamic evolution of the traffic density. In the literature, this version of the CTM has been extended to consider also the dynamics of the queue lengths present at the on-ramps and the possibility to regulate the flow entering from the on-ramp via ramp metering control. This augmented version is adopted especially when ramp metering control approaches are designed, for instance, in [51–53].

In this case, the following dynamic quantities are added to the model (see Fig. 3.9):

- $l_i(k)$ is the queue length of vehicles waiting in the on-ramp of cell i at time kT [veh];
- $d_i(k)$ is the flow accessing the on-ramp of cell i during time interval $[kT, (k + 1)T)$ [veh/h];
- $r_i^C(k)$ is the *ramp metering control variable*, i.e. the flow determined by the ramp metering controller to enter cell i from the on-ramp during time interval $[kT, (k + 1)T)$ [veh/h].

The parameter r_i^{max} is also considered, representing the capacity of the on-ramp of section i , i.e. the maximum flow that can enter from that on-ramp, $i = 1, \dots, N$.

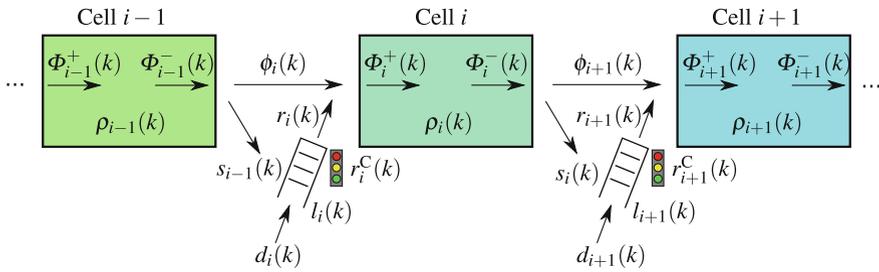


Fig. 3.9 Sketch of freeway stretch in case of on-ramp queues and the relative notation in the CTM

The dynamic equation of the on-ramp queue length, for $i = 1, \dots, N$, $k = 0, \dots, K - 1$, is given by

$$l_i(k + 1) = l_i(k) + T [d_i(k) - r_i(k)] \quad (3.37)$$

In this case, the on-ramp demand $D_i^{\text{ramp}}(k)$ to be used in (3.33) and (3.34) is no more a boundary condition but it is computed by taking into account the queue length evolution and the flow accessing the on-ramp. It is possible to distinguish between two cases, corresponding, respectively, to uncontrolled and controlled on-ramp flows.

Uncontrolled On-Ramps If the on-ramp in section i is not controlled, the on-ramp demand of cell i during time interval $[kT, (k + 1)T)$ is given by

$$D_i^{\text{ramp}}(k) = \min \left\{ d_i(k) + \frac{l_i(k)}{T}, r_i^{\text{max}} \right\} \quad (3.38)$$

Controlled On-Ramps If the on-ramp in section i is controlled, the on-ramp demand of cell i is given by

$$D_i^{\text{ramp}}(k) = \min \left\{ d_i(k) + \frac{l_i(k)}{T}, r_i^C(k), r_i^{\text{max}} \right\} \quad (3.39)$$

The augmented CTM to include the on-ramp queue dynamics, for a freeway with off-ramps and on-ramps in all the cells, is given by (3.27)–(3.34), (3.37), together with (3.38) for the uncontrolled on-ramps and (3.39) for the controlled on-ramps. The *boundary conditions* are now the demand in the cell before the first one, i.e. $D_0(k)$, the supply of the cell after the last one, i.e. $S_{N+1}(k)$, the flows accessing the on-ramp queues, i.e. $d_i(k)$, and the split ratios, i.e. $\beta_i(k)$, $i = 1, \dots, N$, $k = 0, \dots, K$.

3.3.3 The CTM in a Mixed-Integer Linear Form

Another version of the CTM is the reformulation of the model in a mixed-integer linear form, i.e. as a *Mixed Logical Dynamical* (MLD) system. According to the framework proposed in [54], an MLD system is a dynamic system characterised by logic rules, on/off inputs, piecewise linear functions, discrete states, and can be expressed with linear equalities and inequalities in which continuous and binary variables are involved.

The CTM in MLD form has been first proposed in [51, 55], where it has been used as prediction model in Model Predictive Control (MPC) schemes. The advantage of using the CTM in MLD form is related to computational issues, since the non-linearities present in the original model are avoided, resulting in a mixed-integer linear model which is equivalent to the original one. This is obtained by adding some equalities and inequalities, as well as some auxiliary variables, both binary and continuous.

The non-linear relations present in the CTM are the minimum functions in (3.31) and (3.32), as well as the relations (3.33) and (3.34). Let us start from equation (3.31) and let us introduce a binary variable $\delta_{i-1}^D(k)$ such that $[\delta_{i-1}^D(k) = 1]$ iff $[(1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k) \leq q_{i-1}^{\max}]$. Exploiting the transformations of propositional logic in linear inequalities reported in [54], this latter relation can be transformed as

$$\begin{aligned} (1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k) - q_{i-1}^{\max} &\leq D_{i-1}^{\max}(1 - \delta_{i-1}^D(k)) \\ (1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k) - q_{i-1}^{\max} &\geq \varepsilon + (D_{i-1}^{\min} - \varepsilon)\delta_{i-1}^D(k) \end{aligned} \quad (3.40)$$

where ε is a small tolerance, D_{i-1}^{\max} and D_{i-1}^{\min} are the maximum and minimum value of function $(1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k) - q_{i-1}^{\max}$, respectively, i.e. $D_{i-1}^{\max} = v_{i-1}\rho_{i-1}^{\max}$ and $D_{i-1}^{\min} = -q_{i-1}^{\max}$. Now (3.31) can be substituted by the following equation:

$$D_{i-1}(k) = \delta_{i-1}^D(k)[(1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k)] + (1 - \delta_{i-1}^D(k))q_{i-1}^{\max} \quad (3.41)$$

which is still non-linear, since it contains a multiplication between variables. This non-linearity can be overcome by introducing another auxiliary variable $z_{i-1}^D(k)$, such that $z_{i-1}^D(k) = \delta_{i-1}^D(k)\rho_{i-1}(k)$. Then, (3.41) becomes

$$D_{i-1}(k) = (1 - \beta_{i-1}(k))v_{i-1}z_{i-1}^D(k) + (1 - \delta_{i-1}^D(k))q_{i-1}^{\max} \quad (3.42)$$

The definition $z_{i-1}^D(k) = \delta_{i-1}^D(k)\rho_{i-1}(k)$ can be obtained with the following set of inequalities

$$\begin{aligned} R_{i-1}^{\min}\delta_{i-1}^D(k) &\leq z_{i-1}^D(k) \leq R_{i-1}^{\max}\delta_{i-1}^D(k) \\ z_{i-1}^D(k) &\geq \rho_{i-1}(k) - R_{i-1}^{\max}(1 - \delta_{i-1}^D(k)) \\ z_{i-1}^D(k) &\leq \rho_{i-1}(k) - R_{i-1}^{\min}(1 - \delta_{i-1}^D(k)) \end{aligned} \quad (3.43)$$

in which R_{i-1}^{\max} and R_{i-1}^{\min} can be estimated as the maximum and minimum value of function $\rho_{i-1}(k)$, i.e. $R_{i-1}^{\max} = \rho_{i-1}^{\max}$ and $R_{i-1}^{\min} = 0$.

Analogously, it is possible to consider Eq. (3.32), for which it is necessary to introduce a binary variable $\delta_i^S(k)$ with the following meaning: $[\delta_i^S(k) = 1]$ iff $[w_i(\rho_i^{\max} - \rho_i(k)) \leq q_i^{\max}]$. Such relation can be transformed as follows:

$$\begin{aligned} w_i(\rho_i^{\max} - \rho_i(k)) - q_i^{\max} &\leq S_i^{\max}(1 - \delta_i^S(k)) \\ w_i(\rho_i^{\max} - \rho_i(k)) - q_i^{\max} &\geq \varepsilon + (S_i^{\min} - \varepsilon)\delta_i^S(k) \end{aligned} \quad (3.44)$$

where S_i^{\max} and S_i^{\min} are the maximum and minimum value of function $w_i(\rho_i^{\max} - \rho_i(k)) - q_i^{\max}$, respectively, i.e. $S_i^{\max} = w_i\rho_i^{\max}$ and $S_i^{\min} = -q_i^{\max}$. Now (3.32) can be written as

$$S_i(k) = \delta_i^S(k)[w_i(\rho_i^{\max} - \rho_i(k))] + (1 - \delta_i^S(k))q_i^{\max} \quad (3.45)$$

which is still non-linear; to overcome this, another variable $z_i^S(k)$ is defined as $z_i^S(k) = \delta_i^S(k)\rho_i(k)$. Then, (3.45) becomes

$$S_i(k) = \delta_i^S(k)w_i\rho_i^{\max} - w_iz_i^S(k) + (1 - \delta_i^S(k))q_i^{\max} \quad (3.46)$$

The relation $z_i^S(k) = \delta_i^S(k)\rho_i(k)$ can be replaced by the following set of inequalities:

$$\begin{aligned} R_i^{\min}\delta_i^D(k) &\leq z_i^S(k) \leq R_i^{\max}\delta_i^S(k) \\ z_i^S(k) &\geq \rho_i(k) - R_i^{\max}(1 - \delta_i^S(k)) \\ z_i^S(k) &\leq \rho_i(k) - R_i^{\min}(1 - \delta_i^S(k)) \end{aligned} \quad (3.47)$$

The CTM in MLD form considers a simplified version of the merge model, i.e. of relations (3.33), (3.34) and (3.39). In particular, the CTM in MLD form considers the following simplified merge model:

$$\begin{aligned} \text{If } D_{i-1}(k) + r_i^C(k) &\leq S_i(k) \\ \text{then } \phi_i(k) &= D_{i-1}(k) \\ \text{else } \phi_i(k) &= S_i(k) - r_i^C(k) \end{aligned} \quad (3.48)$$

Since (3.48) is non-linear, it is necessary to introduce a binary variable $\delta_i^M(k)$ defined as $[\delta_i^M(k) = 1]$ iff $[D_{i-1}(k) + r_i^C(k) \leq S_i(k)]$, corresponding to the following inequalities:

$$\begin{aligned} D_{i-1}(k) + r_i^C(k) - S_i(k) &\leq M_i^{\max}(1 - \delta_i^M(k)) \\ D_{i-1}(k) + r_i^C(k) - S_i(k) &\geq \varepsilon + (M_i^{\min} - \varepsilon)\delta_i^M(k) \end{aligned} \quad (3.49)$$

where M_i^{\max} and M_i^{\min} are the maximum and minimum value of function $D_{i-1}(k) + r_i^C(k) - S_i(k)$, respectively, i.e. $M_i^{\max} = q_{i-1}^{\max} + r_i^{\max}$ and $M_i^{\min} = -q_i^{\max}$. It is now possible to write (3.48) as

$$\phi_i(k) = \delta_i^M(k)D_{i-1}(k) + (1 - \delta_i^M(k))[S_i(k) - r_i^C(k)] \quad (3.50)$$

which is still non-linear because of the products between variables. Then, other three variables should be defined. First of all, the auxiliary variable $z_i^{M_D}(k)$ is defined as $z_i^{M_D}(k) = \delta_i^M(k)D_{i-1}(k)$ and corresponds to

$$\begin{aligned} M_{D,i}^{\min} \delta_i^M(k) &\leq z_i^{M_D}(k) \leq M_{D,i}^{\max} \delta_i^M(k) \\ z_i^{M_D}(k) &\geq D_{i-1}(k) - M_{D,i}^{\max} (1 - \delta_i^M(k)) \\ z_i^{M_D}(k) &\leq D_{i-1}(k) - M_{D,i}^{\min} (1 - \delta_i^M(k)) \end{aligned} \quad (3.51)$$

in which $M_{D,i}^{\max}$ and $M_{D,i}^{\min}$ are the maximum and minimum value of function $D_{i-1}(k)$, i.e. $M_{D,i}^{\max} = q_{i-1}^{\max}$ and $M_{D,i}^{\min} = 0$.

Then, the auxiliary variable $z_i^{M_S}(k)$ is defined as $z_i^{M_S}(k) = \delta_i^M(k)S_i(k)$ and given by

$$\begin{aligned} M_{S,i}^{\min} \delta_i^M(k) &\leq z_i^{M_S}(k) \leq M_{S,i}^{\max} \delta_i^M(k) \\ z_i^{M_S}(k) &\geq S_i(k) - M_{S,i}^{\max} (1 - \delta_i^M(k)) \\ z_i^{M_S}(k) &\leq S_i(k) - M_{S,i}^{\min} (1 - \delta_i^M(k)) \end{aligned} \quad (3.52)$$

in which $M_{S,i}^{\max}$ and $M_{S,i}^{\min}$ are the maximum and minimum value of function $S_i(k)$, i.e. $M_{S,i}^{\max} = q_i^{\max}$ and $M_{S,i}^{\min} = 0$.

Finally, the auxiliary variable $z_i^{M_R}(k)$ is defined as $z_i^{M_R}(k) = \delta_i^M(k)r_i^C(k)$ and corresponds to

$$\begin{aligned} M_{R,i}^{\min} \delta_i^M(k) &\leq z_i^{M_R}(k) \leq M_{R,i}^{\max} \delta_i^M(k) \\ z_i^{M_R}(k) &\geq r_i(k) - M_{R,i}^{\max} (1 - \delta_i^M(k)) \\ z_i^{M_R}(k) &\leq r_i(k) - M_{R,i}^{\min} (1 - \delta_i^M(k)) \end{aligned} \quad (3.53)$$

in which $M_{R,i}^{\max}$ and $M_{R,i}^{\min}$ can be estimated as the maximum and minimum value of function $r_i(k)$, i.e. $M_{R,i}^{\max} = r_i^{\max}$ and $M_{R,i}^{\min} = 0$.

Then, (3.50) becomes

$$\phi_i(k) = z_i^{M_D}(k) + S_i(k) - r_i^C(k) - z_i^{M_S}(k) + z_i^{M_R}(k) \quad (3.54)$$

Moreover, the following inequalities must be verified:

$$r_i^C(k) \leq r_i^{\max} \quad (3.55)$$

$$r_i^C(k) \leq d_i(k) + \frac{l_i(k)}{T} \quad (3.56)$$

The CTM in MLD form is given by (3.27)–(3.30), (3.40), (3.42)–(3.44), (3.46)–(3.47), (3.49), (3.51)–(3.56). Note that the CTM in MLD form is characterised by three sets of auxiliary binary variables, namely, $\delta_i^D(k)$, $\delta_i^S(k)$, $\delta_i^M(k)$, $i = 1, \dots, N$, $k = 0, \dots, K$, and five sets of auxiliary continuous variables, namely, $z_i^D(k)$, $z_i^S(k)$, $z_i^{M_D}(k)$, $z_i^{M_S}(k)$, $z_i^{M_R}(k)$, $i = 1, \dots, N$, $k = 0, \dots, K$.

3.3.4 The CTM Including Capacity Drop Phenomena

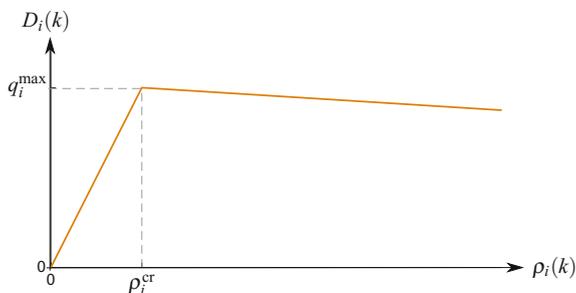
First-order traffic models, both of continuous and discrete type, are not able to capture the capacity drop, which is a common traffic phenomenon detected in real cases (see Sect. 2.2.4). In the literature, some research works have been devoted to include capacity drop phenomena in first-order traffic models (see, e.g. [56–60]). In the following, two interesting extended versions of the CTM are reported, respectively, obtained by changing the demand function and by changing both the demand and the supply according to a 5-step piecewise linear Fundamental Diagram.

CTM with Capacity Drop: Change in the Demand Function A possibility of modelling capacity drop phenomena in the CTM has been proposed in [60, 61]. In this model, the drop is represented by simply modifying the demand function, so that in case of congestion the demand function is linearly decreasing, as shown in Fig. 3.10. More specifically, the demand of cell $i - 1$, for $i = 1, \dots, N$ and $k = 0, \dots, K$, instead of being represented by (3.31), is given by

$$D_{i-1}(k) = \min \left\{ (1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k), q_{i-1}^{\max} + w'_{i-1}(\rho_{i-1}^{\text{cr}} - \rho_{i-1}(k)) \right\} \quad (3.57)$$

where w'_i is the decreasing capacity rate due to the capacity drop phenomenon referred to cell i ($w'_i < w_i$), while ρ_i^{cr} is the critical density of cell i causing a breakdown in capacity.

Fig. 3.10 Modified demand function (CTM with capacity drop)



This simple modification of the demand function allows to model capacity drop by maintaining a simple linear formulation of the model that is useful especially for control purposes. Of course, this simple modification is not sufficient to create a capacity drop at the head of a congestion under all circumstances [60].

A very recent extension of the CTM to include the capacity drop phenomenon has been proposed in [62]. This model has been used in a model-based predictive control scheme in [63], in which it has been extended to consider the application of variable speed limits. In [63], the proposed modified CTM is described in comparison with the one reported in [60]. The model discussed in [63] is validated in [64], where it is calibrated with real traffic data from a Dutch freeway and compared with a second-order traffic flow model.

CTM with Capacity Drop: a 5-step Piecewise Linear Fundamental Diagram

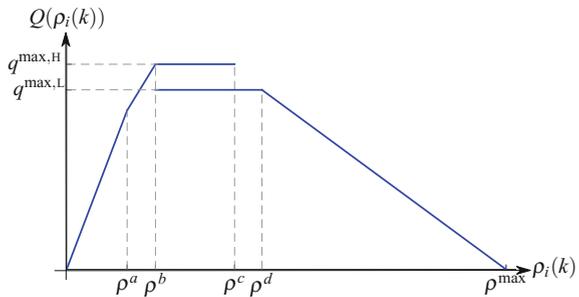
Another possibility of including the capacity drop in the CTM has been proposed in [59]. In that work, a 5-step piecewise linear Fundamental Diagram is defined, based on empirical data, in which two values of capacity are explicitly considered. Then, the capacity drop (between these two values of capacity) is modelled by introducing a memory-state binary variable which determines whether the bottleneck is active or inactive.

Let us start from the 5-step piecewise linear Fundamental Diagram. In a given location the steady-state relation between traffic flow and density is assumed to be a 5-step piecewise linear function, which can be written as follows:

$$Q(\rho_i(k)) = \begin{cases} v\rho_i(k) & \text{if } 0 \leq \rho_i(k) \leq \rho^a \wedge \sigma_i(k) = 0 \\ \kappa + v'\rho_i(k) & \text{if } \rho^a \leq \rho_i(k) \leq \rho^b \wedge \sigma_i(k) = 0 \\ q^{\max,H} & \text{if } \rho^b \leq \rho_i(k) \leq \rho^c \wedge \sigma_i(k) = 0 \\ q^{\max,L} & \text{if } \rho^b \leq \rho_i(k) \leq \rho^d \wedge \sigma_i(k) = 1 \\ w(\rho^{\max} - \rho_i(k)) & \text{if } \rho^d \leq \rho_i(k) \leq \rho^{\max} \wedge \sigma_i(k) = 1 \end{cases} \quad (3.58)$$

A representation of the piecewise linear relation (3.58) is given in Fig. 3.11. Each block of this function is defined by the density boundaries ρ^a , ρ^b , ρ^c , ρ^d , the jam density ρ^{\max} and the congestion state $\sigma_i(k) \in \{0, 1\}$. This latter is a binary quantity,

Fig. 3.11 5-step piecewise linear approximation of the Fundamental Diagram



equal to 0 when the state is uncongested and equal to 1 when it is congested. Note that the density boundaries must verify $0 < \rho^a < \rho^b < \rho^c < \rho^d < \rho^{\max}$.

The first two blocks represent the uncongested phase of traffic. The first block is for *light conditions* in which vehicles move at free-flow speed v . The second block represents the *undersaturated* state of traffic, in which the interactions among vehicles decrease the mean speed (that is equal to $v' < v$). The third and fourth blocks represent, respectively, the *pre-congestion* and *post-congestion* situations. Indeed, there is a time interval in which, despite the high density, the freeway works at the maximum capacity $q^{\max,H}$. After that time, a breakdown occurs and capacity decreases to a lower value, that is $q^{\max,L}$. Finally, the fifth block represents the behaviour in the *congested* phase; therefore, the (negative) slope is assumed to be equal to the congestion wave speed w . Moreover, if the density value is equal to the maximum value ρ^{\max} , the flow is equal to zero.

According to this 5-step piecewise linear Fundamental Diagram, the standard CTM has been modified in [59], by changing the demand and supply functions, as well as by introducing a relation to update the value of the congestion state variable. Further parameters are added to the standard ones in the CTM. Such parameters, referred to cell i , $i = 1, \dots, N$, are the undersaturated speed v'_i [km/h], the constant κ_i [veh/h], the high and low capacity values $q_i^{\max,H}$ and $q_i^{\max,L}$ [veh/h], the density boundaries $\rho_i^a, \rho_i^b, \rho_i^c, \rho_i^d$ [veh/km].

Taking into account (3.58) and Fig. 3.11, it is possible to split the graph into two parts: the left part of the graph (from the first to the third block) is related to the demand function, while the right part (fourth and fifth blocks) is associated with the supply function. Specifically, the demand of cell $i - 1$ and the supply of cell i , instead of being given by (3.31) and (3.32), are, respectively, defined as

$$D_{i-1}(k) = \min \left\{ (1 - \beta_{i-1}(k))v_{i-1}\rho_{i-1}(k), (1 - \beta_{i-1}(k))[\kappa_{i-1} + v'_{i-1}\rho_{i-1}(k)], q_{i-1}^{\max,H} \right\} \quad (3.59)$$

$$S_i(k) = \begin{cases} \min \left\{ w_i(\rho_i^{\max} - \rho_i(k)), q_i^{\max,H} \right\} & \text{if } \sigma_i(k-1) = 0 \\ \min \left\{ w_i(\rho_i^{\max} - \rho_i(k)), q_i^{\max,L} \right\} & \text{if } \sigma_i(k-1) = 1 \end{cases} \quad (3.60)$$

The congested state variable $\sigma_i(k)$ indicates if the state of cell i at time kT is uncongested or congested and is given by

$$\sigma_i(k) = \begin{cases} 1 & \text{if } (\rho_i(k) \geq \rho_i^c) \vee (\rho_i(k) \geq \rho_i^b \wedge \sigma_i(k-1) = 1) \\ 0 & \text{otherwise} \end{cases} \quad (3.61)$$

According to [59], in the 5-step piecewise linear Fundamental Diagram, there is not one value for the critical density, but two values, i.e. ρ_i^b and ρ_i^c . These values of density are responsible for changing $\sigma_i(k)$ to 0 or to 1.

3.3.5 The CTM for a Freeway Network

The CTM for a freeway network has been first introduced in [45], in which only three-legged junctions are modelled. According to this assumption, the cells are classified into three types (see Fig. 3.12):

- a *diverge* cell is characterised by only one entering link and two leaving links;
- a *merge* cell presents two entering links and one exiting link;
- an *ordinary* cell has just one entering and one leaving link.

With these three types of cells, any freeway networks with three-legged junctions can be modelled. Nevertheless, no generality is lost because the case of junctions with more than three legs can be easily represented as combinations of three-legged junctions, as discussed in [45].

The dynamics of ordinary cells has already been described in Sect. 3.3.1. As for merge and diverge cells, the state equation for traffic density (3.27) still holds, but the definition of entering and exiting flows should be modified. In particular, for a merge cell i , the total flow $\Phi_i^+(k)$ entering cell i during time interval $[kT, (k+1)T)$ depends on the flows coming from the preceding cells. Conversely, for a diverge cell i , the total flow $\Phi_i^-(k)$ exiting cell i during time interval $[kT, (k+1)T)$ depends on the flows going to the following cells. Let us analyse these two cases separately.

Merge Cell Let us consider that cell i is a merge cell and let us denote with j and l the two preceding cells. Let us denote with $\phi_{j,i}(k)$ and $\phi_{l,i}(k)$ the flows entering cell i during time interval $[kT, (k+1)T)$, from cell j and l , respectively, as shown in Fig. 3.13.

For cells j and l it is possible to define the demand, i.e. the flow that can be sent from cell j and l , respectively, to cell i during time interval $[kT, (k+1)T)$. Analogously to (3.31), these demands can be computed as

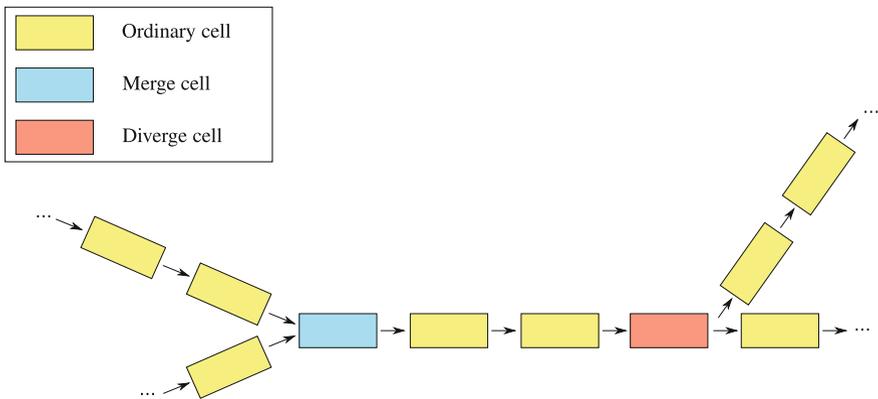
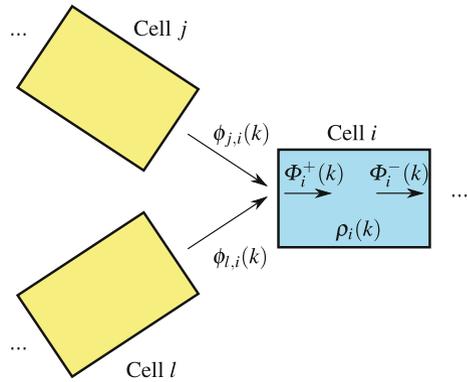


Fig. 3.12 Different types of cells in a freeway network according to the CTM

Fig. 3.13 Sketch of a merge cell and the relative notation



$$D_j(k) = \min \{ (1 - \beta_j(k)) v_j \rho_j(k), q_j^{\max} \} \quad (3.62)$$

$$D_l(k) = \min \{ (1 - \beta_l(k)) v_l \rho_l(k), q_l^{\max} \} \quad (3.63)$$

The supply of cell i represents the flow that can be received by cell i from cells j and l during time interval $[kT, (k+1)T)$ and is still given by (3.32).

The total flow $\Phi_i^+(k)$ entering cell i is computed as the sum of the flows entering from cells j and l , i.e.

$$\Phi_i^+(k) = \phi_{j,i}(k) + \phi_{l,i}(k) \quad (3.64)$$

As already analysed in Sect. 3.3.1 for the mainstream flow and the on-ramp flow, we are again in the situation in which there are two sending cells and one receiving cell. Two cases are distinguished, corresponding to free-flow and congested conditions. In the free-flow case, in cell i there is enough space for the two flows coming from cells j and l , and a condition analogous to (3.33) can be written, i.e.

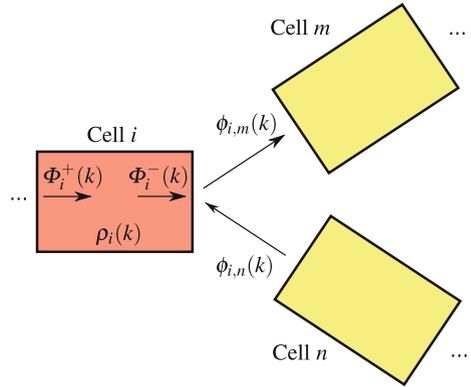
$$\begin{aligned} \text{If } D_j(k) + D_l(k) &\leq S_i(k) \\ \text{then } \phi_{j,i}(k) &= D_j(k), \quad \phi_{l,i}(k) = D_l(k) \end{aligned} \quad (3.65)$$

If the previous condition is not satisfied, this means that not all the flows coming from cells j and l can be received by cell i , and, analogously to (3.34), the following conditions for the congested case hold:

$$\begin{aligned} \text{If } D_j(k) + D_l(k) &> S_i(k) \\ \text{then } \phi_{j,i}(k) &= \text{mid} \{ D_j(k), S_i(k) - D_l(k), p_j S_i(k) \} \\ \phi_{l,i}(k) &= \text{mid} \{ D_l(k), S_i(k) - D_j(k), p_l S_i(k) \} \end{aligned} \quad (3.66)$$

where p_j and p_l are the priorities of cell j and l in the merge, with $p_j + p_l = 1$.

Fig. 3.14 Sketch of a diverge cell and the relative notation



Diverge Cell Let us consider that cell i is a diverge cell and let us denote with m and n the two following cells. Let us denote with $\phi_{i,m}(k)$ and $\phi_{i,n}(k)$ the flows exiting cell i during time interval $[kT, (k+1)T)$ and going to cells m and n , respectively, as shown in Fig. 3.14.

For cell i , the demand $D_i(k)$ is the flow that can be sent from cell i to cells m and n during time interval $[kT, (k+1)T)$, respectively, and it is computed as in (3.31), i.e.

$$D_i(k) = \min \{ (1 - \beta_i(k))v_i\rho_i(k), q_i^{\max} \} \quad (3.67)$$

The supply of cells m and n is instead the flow that can be received by cells m and n , respectively, from cell i during time interval $[kT, (k+1)T)$. These quantities are computed analogously to (3.32), i.e.

$$S_m(k) = \min \{ w_m(\rho_m^{\max} - \rho_m(k)), q_m^{\max} \} \quad (3.68)$$

$$S_n(k) = \min \{ w_n(\rho_n^{\max} - \rho_n(k)), q_n^{\max} \} \quad (3.69)$$

The total flow $\Phi_i^-(k)$ exiting cell i is computed by taking into account the assumptions of the diverge model. The basic idea is that this total flow is restricted in case at least one of the two diverging branches cannot receive its allocated flow. According to this assumption, vehicles which cannot go to the next cell prevent all the other vehicles behind them to continue, supposing that vehicles at the diverge area are served according to a first-in-first-out rule. Of course, this is not completely true in real cases, especially for low exit percentages, but it is worth noting that some blockage phenomena can occur in reality for high exit percentages and in specific traffic conditions, in some way motivating this assumption.

By denoting with $\beta_m(k)$ and $\beta_n(k)$ the portions of traffic flow present in cell i going, respectively, to cell m and n (supposing that these quantities are exogenously determined), the total flow exiting cell i is computed as

$$\Phi_i^-(k) = \min \left\{ D_i(k), \frac{S_m(k)}{\beta_m(k)}, \frac{S_n(k)}{\beta_n(k)} \right\} \quad (3.70)$$

The flows exiting cell i and going to cells m and n are then computed as

$$\phi_{i,m}(k) = \beta_m(k) \Phi_i^-(k) \quad (3.71)$$

$$\phi_{i,n}(k) = \beta_n(k) \Phi_i^-(k) \quad (3.72)$$

3.3.6 Other CTM Versions

Other modifications of the original CTM have been proposed in the literature in the last two decades. Hereafter some of them are briefly commented for the reader's convenience, while others can be addressed with the relevant references in [65]. Note that some of the modifications regard the extension of the CTM to include the case of a freeway in which variable speed limits or route guidance strategies are applied. These modifications to the CTM are not reported in this book, whereas these types of control have been included in second-order models (see Chap. 4), being this latter the most common choice in the scientific literature.

Asymmetric Cell Transmission Model The Asymmetric Cell Transmission Model (ACTM) is a modification of the CTM proposed in [52]. The relevant difference between the two models is the treatment of traffic merges. More specifically, merges in the ACTM are considered as asymmetric connections, such as the junctions of the on-ramps into the mainstream. According to the logic of the standard CTM, the merge is oriented to move as much of the demand as possible from the two merging cells into the receiving cell. The ACTM, instead, makes separate allocations of supply for each merging flow. The flows can then be computed separately as the minimum among the demand, the allocated supply, and the capacity. This modification is justified by the fact that the non-concave/non-convex mid functions of the CTM in (3.34) are replaced with concave min functions, which is an advantage when this model is used as a basis to solve model-based traffic control problems. Moreover, in [52] it is proved that the ACTM, as the CTM, ensures not to predict unrealistic behaviours such as backward moving traffic, negative densities and densities exceeding the jam density.

Link-Node Cell Transmission Model The Link-Node Cell Transmission Model (LN-CTM) is an extension of the CTM to simulate traffic in road networks [66]. In this model, the traffic network is represented with a directed graph, in which links represent road segments and nodes are the junctions among links. Normal links are used to connect two nodes, source links are used to introduce traffic in the network, whereas sink links are used to receive traffic moving out of the network. According to this logic, the on-ramps are represented as source links, while the off-ramps are sinks. The LN-CTM uses a more accurate model of the merging phenomena compared

with the ACTM. In particular, in congested conditions, the available supply is shared by the incoming flows proportionally to the demands. However, this more detailed representation of the merge comes at an additional cost of added non-linearity [67], and therefore the results proved in [52] cannot be applied for the LN-CTM.

Lagged Cell Transmission Model In order to improve the accuracy of the original CTM, in [68] a modification of the model has been proposed, based on the fact that the downstream density, used to calculate the supply, is measured at an earlier time instant compared to the current time step, i.e. it is lagged. The introduction of lags can be justified because traffic information travels more slowly in the upstream than in the downstream direction. An improved version of the Lagged CTM has been proposed in [69] to avoid the occurrence of negative densities and of densities larger than the maximum value.

Variable-Length Cell Transmission Model The Variable-Length Cell Transmission Model (VLM) has been proposed in [70] and differs from the standard CTM for the fact that a limited number of cells (of variable length) are used. A road network is subdivided into several sections which are assumed to be composed of a downstream congested cell followed by a free upstream cell. Both cells have variable lengths and are described by two lumped densities (one congested, the other free). The model includes one more state describing the length variation for each cell.

Switched Interpretation of CTM In the literature, some switched interpretations of the CTM have also appeared. Indeed, the CTM is a piecewise linear model and can be regarded as a hybrid system that switches among different sets of linear difference equations. Each set describes a specific operation mode of the freeway traffic system. Since the number of modes can become very high [71, 72], some assumptions can be made to reduce the number of modes. A typical assumption is to consider at most one wave front in the considered freeway stretch. The presence of a single wave front is an assumption reasonable for short freeway stretches with only one on-ramp and one off-ramp. The switched interpretation of the CTM with the single-wave front assumption is called in the literature *Switching-Mode Model (SMM)* [73]. The reduced set of modes can be associated with a graph, since the transition between modes has to follow specific rules, also dictated by the fact that the congestion moves upwards or downwards. The switched model with the associated graph is regarded as a *Graph Constrained CTM* [74].

3.4 Multi-class First-Order Models

Multi-class traffic models have been developed by researchers in order to distinguish different classes of vehicles travelling in the same road system. Depending on the objective of the model, the vehicle classes can be referred to different types of vehicles (e.g. cars, trucks, public transport vehicles and so on) or to specific features of the drivers (such as driving behaviours, travel purposes and so on). In recent applications,

it is becoming more and more relevant to distinguish vehicles according to the driver information level, specifically representing the class of ‘intelligent vehicles’, i.e. vehicles equipped with innovative technology enabling the exchange of data with other vehicles and the traffic infrastructure.

3.4.1 Motivations for Multi-class Models

Regardless of the considered vehicle typologies, multi-class models are characterised by a higher descriptive capability than single-class models, allowing to more realistically represent the dynamic behaviour of a real traffic system. Multi-class models may allow the description of relevant traffic phenomena that can not be captured by models representing only one class of vehicles, in particular all the phenomena related to the interaction of different groups of vehicles which have to share the same infrastructure.

Referring specifically to macroscopic traffic models, a multi-class macroscopic model assumes that the traffic behaviour is represented as the interaction of different traffic flows corresponding to different vehicle categories, whereas a single-class model assumes that the whole traffic is a homogeneous fluid. Let us consider in particular the easiest and better known example of multi-class traffic, i.e. a freeway traffic system in which both cars and trucks travel. In this case, it is easy to observe that trucks have a strong impact on the overall traffic flow for many reasons (because of their high dimensions and low operating capabilities, because their presence has a psychological impact on the drivers of nearby vehicles and so on). Also, these two classes of vehicles have different behaviours and, in many traffic scenarios, can be seen as two different flows sharing the same infrastructure. In particular considering roads with multiple lanes, as it normally happens in freeways, fast vehicles can overtake slow vehicles, so that the traffic behaviour is given by the dynamics of two different flows which influence each other. Representing explicitly the two flows and the interaction between them, instead of modelling the whole traffic as a unique flow, allows to better describe the real traffic system.

A further advantage of the multi-class modelling framework is related to the possibility of designing multi-class controllers, providing different control actions for different vehicle classes. This aspect will be further investigated in Chap. 10.

3.4.2 An Overview of Multi-class First-Order Models

Most of the multi-class first-order traffic models present in the literature are multi-class versions of the LWR model, while only few are multi-class extensions of the CTM.

Multi-class Versions of the LWR Model In some cases, the heterogeneous properties of the traffic flow are represented through *multi-lane* models, as in [75], where a general multi-lane rule is introduced, or in [76], where two types of vehicles and a set of dedicated lanes are modelled. In particular, in [76], the vehicles of the first class can use all the lanes, whereas those belonging to the second class are compelled to travel in a subset of lanes usually located on the right side of the freeway.

Another multi-class first-order model is reported in [77], where the macroscopic model is derived from mesoscopic principles, i.e. from gas-kinetic equations, in order to model the fact that drivers accelerate/decelerate not only according to the desired speed of their class but also due to interactions with other vehicles, both belonging to their class and to other classes. The model proposed in [77] is extended in [78], where a multi-class multi-lane model is proposed for explicitly representing the presence of vehicles moving in platoons. A macroscopic behavioural theory of traffic dynamics for homogeneous and multi-lane freeways is proposed in [79, 80]. Taking into account that drivers can be distinguished in timid and aggressive, this theory can be used to make predictions for separate groups of lanes and is shown to be consistent with experimental observations.

A kinematic wave model of multi-commodity network traffic flow is presented in [81] and, then, extended in [82] to the lane-changing case. In these works, it is assumed that all vehicles have predefined paths and each commodity is represented by vehicles using the same path.

Another work developed in order to consider heterogeneous groups of drivers in the traffic flow is [83], where an extension of the LWR model is formulated with different speed distributions for each class of road users. Specifically, that model describes the dynamic behaviour of heterogeneous users in the traffic flow, in which faster vehicles can overtake the slower ones, both under uncongested and congested conditions, whereas slower vehicles behave in order to slow down the faster ones. In [84], the authors present a homogenised hyperbolic traffic flow model to take into account the presence of several types of vehicles (such as cars, trucks, buses, and so on). An n -population generalisation of the LWR model is proposed in [85], allowing to mathematically explain some practical traffic phenomena, such as overtaking among vehicles. In [86], a multi-class first-order model is presented to explain non-linear traffic phenomena, such as hysteresis and capacity drop. In that model, in free-flow conditions the different vehicle classes are characterised by specific desired speeds and overtaking is allowed; in congested conditions, instead, all the vehicles must travel at the same congested speed and it is not possible to overtake.

In [87], a new model is proposed for vehicle classes interacting in a non-cooperative way: slow vehicles can be seen as *moving bottlenecks* for the fast vehicles, which maximise their speed without influencing slower vehicles. In this model, each class represents a homogeneous group of vehicles, which interacts with the other vehicle classes within the traffic flow. According to this concept, each class is characterised by a different Fundamental Diagram. Note that the specific case of moving bottlenecks, i.e. the presence of slow vehicles moving in the traffic flow, resulting in a reduction of the capacity, is studied in a number of research works based on the LWR model (see, e.g. [88–92]). From a mathematical point of view, moving bottlenecks

are normally represented as models in which the partial differential equations of the LWR model are coupled with ordinary differential equations describing the motion of slower vehicles.

Based on the same logic applied in [87], a model to describe a *disordered traffic system* is presented in [93]. In a disordered traffic system, there is no lane discipline, i.e. drivers of smaller vehicles exploit their manoeuvrability to move into lateral gaps at low speeds, whereas at high speeds larger vehicles exploit their greater power to move forward in the traffic flow. These types of systems are very common in developing countries.

A more recent development of macroscopic first-order models for the multi-class case is the Fastlane model, which was first developed in [94]. Fastlane was then successively extended in [95] to be applied for developing multi-class ramp metering in order to control separately the different vehicle classes. Fastlane is based on the LWR model and differs from earlier multi-class first-order macroscopic traffic models for the fact that it models the system dynamics in terms of state-dependent (instead of constant) passenger car equivalents. According to the Highway Capacity Manual, the *Passenger Car Equivalents* (PCE) are defined as the number of passenger cars displaced by a single heavy vehicle of a particular type under prevailing roadway, traffic and control conditions [96]. This factor depends on the considered freeway portion and the traffic conditions present in it, as discussed, for instance, in [97].

A recent work on multi-class traffic models is [98], where two types of vehicles are considered. The model is able to capture overtaking dynamics and creeping phenomena, these latter representing overtaking actions by small vehicles in highly congested situations when larger vehicles have completely stopped. In [98], it is shown that this two-class homogeneous model is equivalent to the ARZ model, of second-order type (see Sect. 4.1.2).

Multi-class Versions of the CTM In [99], the conventional single-class CTM is extended to a more generalised multi-class model in order to take into account the mixed composition of vehicle classes. In the experimental results reported in the paper, the multi-class model is compared with the single-class one and is proven to be significantly more accurate in representing real traffic scenarios.

In [100], a multi-class CTM is developed with the aim of distinguishing two specific classes of vehicles, i.e. *autonomous vehicles* and conventional vehicles. Indeed, autonomous vehicles may entail reduced headways and an increased capacity. Of course, this impact depends on the proportion of autonomous vehicles in the entire traffic flow. The idea of explicitly modelling the presence of Vehicle Automation and Communication Systems (VACS) in the traffic network can be also found in [60], where the CTM is modified to consider lane-changing and capacity drop phenomena, by specifically computing lateral and longitudinal flows.

A recent multi-class version of the CTM can be found in [101], where a unified framework to model heterogeneous traffic flows for large-scale networks is proposed. This model considers the interaction of different vehicle classes, each of which is characterised by homogeneous car-following behaviours and vehicle attributes, and represents three traffic states, corresponding, respectively, to free-flow, semi-

congested, and full congested conditions. This model also allows the computation of travel times for each vehicle class.

A multi-class version of the CTM, specifically modelling the presence of cars and buses in the traffic flow, is presented in [102]. The proposed model is called BUS-CTM and tries to replicate the phenomenon of moving bottlenecks, caused by buses moving in the traffic flow. Specifically, buses and cars are considered as heterogeneous vehicles with different characteristics, such as free-flow speed, acceleration and size.

References

1. Piccoli B, Tosin A (2009) Vehicular traffic: a review of continuum mathematical models. In: Meyers RA (ed) Encyclopedia of complexity and systems science. Springer, Berlin, pp 9727–9749
2. Greenshields BD, Bibbins JR, Channing WS, Miller HH (1935) A study of traffic capacity. In: Highway research board proceedings, vol 14, pp 448–477
3. Papageorgiou M, Blosseville J-M, Hadj-Salem H (1989) Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. *Transp Res Part B* 23:29–47
4. Lighthill MJ, Whitham GB (1955) On kinematic waves II: a theory of traffic flow on long crowded roads. *Proc R Soc A* 229:317–345
5. Richards PI (1956) Shock waves on the highway. *Oper Res* 4:42–51
6. Lax PD (1973) Hyperbolic systems of conservation laws and the mathematical theory of shock waves. SIAM, Philadelphia
7. Leveque RJ (1992) Numerical methods for conservation laws. Birkhäuser Verlag, Basel
8. Evans L (1998) Partial differential equations. American Mathematical Society, Providence
9. Serre D (1996) Systems of conservation laws. Cambridge University Press, Cambridge
10. Bressan A (2000) Hyperbolic systems of conservation laws. Oxford University Press, Oxford
11. Garavello M, Piccoli B (2016) Traffic flow on networks. American Institute of Mathematical Sciences
12. Garavello M, Han K, Piccoli B (2016) Models for vehicular traffic on networks. American Institute of Mathematical Sciences
13. Papageorgiou M (1998) Some remarks on macroscopic traffic flow modelling. *Transp Res Part A* 32:323–329
14. Lax P (1953) Nonlinear hyperbolic equations. *Commun Pure Appl Math* 6:231–258
15. Glimm J (1965) Solutions in the large for nonlinear hyperbolic systems of equations. *Commun Pure Appl Math* 18:697–715
16. Kruzkov S (1970) First order quasilinear equations in several independent variables. *Math USSR-Sbornik* 10:217–243
17. DiPerna RJ (1976) Global existence of solutions to nonlinear hyperbolic systems of conservation laws. *J Differ Equ* 20:187–212
18. Bardos C, Leroux A, Nedelec J (1979) First order quasilinear equations with boundary conditions. *Commun Part Differ Equ* 4:1017–1034
19. Lebacque J (1996) The Godunov scheme and what it means for first order traffic flow models. In: Proceedings of the 13th international symposium on transportation and traffic theory, pp 647–677
20. Li T (2003) Global solutions of nonconcave hyperbolic conservation laws with relaxation arising from traffic flow. *J Differ Equ* 190:131–149
21. Li T, Liu H (2005) Stability of a traffic flow model with nonconvex relaxation. *Commun Math Sci* 3:101–118

22. Dubois F, Le Floch P (1988) Boundary conditions for nonlinear hyperbolic systems of conservation laws. *J Differ Equ* 71:93–122
23. Le Floch P (1988) Explicit formula for scalar non-linear conservation laws with boundary condition. *Math Methods Appl Sci* 10:265–287
24. Strub IS, Bayen AM (2006) Weak formulation of boundary conditions for scalar conservation laws: an application to highway traffic modelling. *Int J Robust Nonlinear Control* 16:733–748
25. Colombo RM, Goatin P, Rosini MD (2011) On the modelling and management of traffic. *ESAIM Math Model Numer Anal* 45:853–872
26. Bagnerini P, Colombo RM, Corli A (2006) On the role of source terms in continuum traffic flow models. *Math Comput Model* 44:917–930
27. Jin W-L, Zhang HM (2003) The inhomogeneous kinematic wave traffic flow model as a resonant nonlinear system. *Transp Sci* 37:294–311
28. Li J, Zhang HM (2013) Modeling space-time inhomogeneities with the kinematic wave theory. *Transp Res Part B* 54:113–125
29. Claudel CG, Bayen AM (2010) Lax-Hopf based incorporation of internal boundary conditions into Hamilton–Jacobi equation, part I: theory. *IEEE Trans Autom Control* 55:1142–1157
30. Claudel CG, Bayen AM (2010) Lax-Hopf based incorporation of internal boundary conditions into Hamilton–Jacobi equation, part II: computational methods. *IEEE Trans Autom Control* 55:1158–1174
31. Holden H, Risebro NH (1995) A mathematical model of traffic flow on a network of unidirectional roads. *SIAM J Math Anal* 26:999–1017
32. Coclite GM, Garavello M, Piccoli B (2005) Traffic flow on a road network. *SIAM J Math Anal* 36:1862–1886
33. Bretti G, Natalini R, Piccoli B (2006) Numerical approximations of a traffic flow model on networks. *Netw Heterog Media* 1:57–84
34. Helbing D, Lämmer S, Lebacque J-P (2005) Self-organized control of irregular or perturbed network traffic. In: Deissenberg C, Hartl RF (eds) *Opt Control Dyn Games*. Springer, Dordrecht, pp 239–274
35. Marigo A, Piccoli B (2008) A fluid dynamic model for T-junctions. *SIAM J Math Anal* 39:2016–2032
36. Delle Monache ML, Reilly J, Samaranayake S, Krichene W, Goatin P, Bayen AM (2014) A PDE-ODE model for a junction with ramp buffer. *SIAM J Appl Math* 74:22–39
37. Herty M, Lebacque J-P, Moutari S (2009) A novel model for intersections of vehicular traffic flow. *Netw Heterog Media* 4:813–826
38. Garavello M, Goatin P (2012) The Cauchy problem at a node with buffer, discrete and continuous dynamical systems, Series A. American Institute of Mathematical Sciences
39. Garavello M, Piccoli B (2013) A multibuffer model for LWR road networks. In: Ukkusuri S, Ozbay K (eds) *Advances in dynamic network modeling in complex transportation systems*. Springer, New York, pp 143–161
40. Leveque RJ (2002) *Finite volume methods for hyperbolic problems*. Cambridge University Press, Cambridge
41. Caligaris C, Sacone S, Siri S (2011) On an implicit and stable resolution scheme for the Payne–Whitham model. *Math Comput Model* 54:378–387
42. Trefethen LN (1996) *Finite difference and spectral methods for ordinary and partial differential equations*. Cornell University
43. Daganzo CF (1993) The cell transmission model part I: a simple dynamic representation of highway traffic. University of California, Berkeley
44. Daganzo CF (1994) The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transp Res Part B* 28:269–287
45. Daganzo CF (1995) The cell transmission model, part II: network traffic. *Transp Res Part B* 29:79–93
46. Daganzo CF (1995) A finite difference approximation of the kinematic wave model of traffic flow. *Transp Res Part B* 29:261–276

47. Godunov SK (1959) A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. *Matematicheskii Sbornik* 47:271–306
48. Yperman I, Logghe S, Immers B (2005) The link transmission model: an efficient implementation of the kinematic wave theory in traffic networks. *Advanced OR and AI methods in transportation*, pp 122–127
49. Newell GF (1993) A simplified theory of kinematic waves in highway traffic. *Transp Res Part B* 27:281–313
50. Himpe W, Corthout R, Tampère MJC (2016) An efficient iterative link transmission model. *Transp Res Part B* 92:170–190
51. Ferrara A, Sacone S, Siri S (2015) Event-triggered model predictive schemes for freeway traffic control. *Transp Res Part C* 58:554–567
52. Gomes G, Horowitz R (2006) Optimal freeway ramp metering using the asymmetric cell transmission model. *Transp Res Part C* 14:244–262
53. Gomes G, Horowitz R, Kurzhanskiy AA, Varaiya P, Kwon J (2008) Behavior of the cell transmission model and effectiveness of ramp metering. *Transp Res Part C* 16:485–513
54. Bemporad A, Morari M (1999) Control of systems integrating logic, dynamics, and constraints. *Automatica* 35:407–427
55. Ferrara A, Nai Oleari A, Sacone S, Siri S (2015) Freeways as systems of systems: a distributed model predictive control scheme. *IEEE Syst J* 9:312–323
56. Lebacque J (1852) Two-phase bounded-acceleration traffic flow model: analytical solutions and applications. *Transp Res Rec* 2003:220–230
57. Laval JA, Daganzo CF (2006) Lane-changing in traffic streams. *Transp Res Part B* 40:251–264
58. Leclercq L, Laval JA, Chiabaut N (2011) Capacity drops at merges: an endogenous model. *Procedia Soc Behav Sci* 17:12–26
59. Srivastava A, Geroliminis N (2013) Empirical observations of capacity drop in freeway merges with ramp control and integration in a first-order model. *Transp Res Part C* 30:161–177
60. Roncoli C, Papageorgiou M, Papamichail I (2015) Traffic flow optimisation in presence of vehicle automation and communication systems - part I: a first-order multi-lane model for motorway traffic. *Transp Res Part C* 57:241–259
61. Roncoli C, Papageorgiou M, Papamichail I (2014) Optimal control for multi-lane motorways in presence of vehicle automation and communication systems. In: *Proceedings of the 19th IFAC world congress*, pp 4178–4183
62. Han Y, Yuan Y, Hegyi A, Hoogendoorn S (2016) New extended discrete first-order model to reproduce propagation of jam waves. *Transp Res Rec* 2560:108–118
63. Han Y, Hegyi A, Yuan Y, Hoogendoorn S, Papageorgiou M, Roncoli C (2017) Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits. *Transp Res Part C* 77:405–420
64. Han Y, Hegyi A, Yuan Y, Hoogendoorn S (2017) Validation of an extended discrete first-order model with variable speed limits. *Transp Res Part C* 83:1–17
65. Alecsandru C, Quddus A, Huang KC, Rouhieh B, Khan AR, Zeng Q (2011) An assessment of the cell-transmission traffic flow paradigm: development and applications. *Transportation Research Board*
66. Muralidharan A, Dervisoglu G, Horowitz R (2009) Freeway traffic flow simulation using the link node cell transmission model. In: *Proceedings of the American control conference*, pp 2916–2921
67. Muralidharan A, Horowitz R (2012) Optimal control of freeway networks based on the link node cell transmission model. In: *Proceedings of the American control conference*, pp 5769–5774
68. Daganzo CF (1999) The lagged cell-transmission model. In: *Proceedings of international symposium on transportation and traffic theory*, pp 81–104
69. Szeto WY (2008) Enhanced lagged cell-transmission model for dynamic traffic assignment. *Transp Res Rec* 76–85
70. Canudas-de-Wit C, Ferrara A (2016) A variable-length cell road traffic model: application to ring road speed limit optimization. In: *Proceedings of the IEEE 55th conference on decision and control*, pp 6745–6752

71. Ferrara A, Sacone S, Siri S, Vivas C, Rubio FR (2016) Switched observer-based ramp metering controllers for freeway systems. In: Proceedings of the 55th IEEE conference on decision and control, pp 6777–6782
72. Ferrara A, Sacone S, Siri S (2015) A switched ramp-metering controller for freeway traffic systems. In: Proceedings of the 5th IFAC conference on analysis and design of hybrid systems, pp 105–110
73. Muñoz L, Sun X, Horowitz R, Alvarez L (2003) Traffic density estimation with the cell transmission model. In: Proceedings of the American control conference, pp 3750–3755
74. Morbidi F, Leon Ojeda L, Canudas de Wit C, Bellicot I (2014) A new robust approach for highway traffic density estimation. In: Proceedings of the European control conference, pp 2575–2580
75. Holland EN, Woods AW (1997) A continuum model for the dispersion of traffic on two-lane roads. *Transp Res Part B* 31:473–485
76. Daganzo CF (1997) A continuum theory of traffic dynamics for freeways with special lanes. *Transp Res Part B* 31:83–102
77. Hoogendoorn SP, Bovy PHL (2000) Continuum modeling of multiclass traffic flow. *Transp Res Part B* 34:123–146
78. Hoogendoorn SP, Bovy PHL (2001) Platoon-based multiclass modeling of multilane traffic flow. *Netw Spat Econ* 1:137–166
79. Daganzo CF (2002) A behavioral theory of multi-lane traffic flow part I: long homogeneous freeway sections. *Transp Res Part B* 36:131–158
80. Daganzo CF (2002) A behavioral theory of multi-lane traffic flow, part II: merges and the onset of congestion. *Transp Res Part B* 36:159–169
81. Jin W-L (2012) A kinematic wave theory of multi-commodity network traffic flow. *Transp Res Part B* 46:1000–1022
82. Jin W-L (2013) A multi-commodity Lighthill–Whitham–Richards model of lane-changing traffic flow. *Transp Res Part B* 57:361–377
83. Wong GCK, Wong SC (2002) A multi-class traffic flow model - an extension of LWR model with heterogeneous drivers. *Transp Res Part A* 36:827–841
84. Bagnerini P, Rascole M (2003) A multiclass homogenized hyperbolic model of traffic flow. *SIAM J Math Anal* 35:949–973
85. Benzoni-Gavage S, Colombo RM (2003) An n -populations model for traffic flow. *Eur J Appl Math* 14:587–612
86. Ngoduy D, Liu R (2007) Multiclass first-order simulation model to explain non-linear traffic phenomena. *Phys A* 385:667–682
87. Logghe S, Immers LH (2008) Multi-class kinematic wave theory of traffic flow. *Transp Res Part B* 42:523–541
88. Lebacque JP, Lesort JB, Giorgi F (1998) Introducing buses into first-order macroscopic traffic flow models. *Transp Res Rec* 1644:70–79
89. Leclercq L (2007) Hybrid approaches to the solutions of the “Lighthill–Whitham–Richards” model. *Transp Res Part B* 41:701–709
90. Lattanzio C, Maurizi A, Piccoli B (2011) Moving bottlenecks in car traffic flow: a PDE-ODE coupled model. *SIAM J Math Anal* 43:50–67
91. Delle Monache ML, Goatin P (2014) Scalar conservation laws with moving constraints arising in traffic flow modeling: an existence result. *J Differ Equ* 257:4015–4029
92. Simoni MD, Claudel CG (2017) A fast simulation algorithm for multiple moving bottlenecks and applications in urban freight traffic management. *Transp Res Part B* 104:238–255
93. Nair R, Mahmassani HS, Miller-Hooks E (2011) A porous flow approach to modeling heterogeneous traffic in disordered systems. *Transp Res Part B* 45:1331–1345
94. van Lint JWC, Hoogendoorn SP, Schreuder M (2008) Fastlane: new multiclass first-order traffic flow model. *Transp Res Rec* 2088:177–187
95. Schreiter T, van Lint H, Hoogendoorn S (2011) Multi-class ramp metering: concepts and initial results. In: Proceedings of the 14th international IEEE conference on intelligent transportation systems, pp 885–889

96. Special Report 209: highway capacity manual. (1994) 3rd edn. Transportation Research Board, Washington DC
97. Al-Kaisy AF, Hall FL, Reisman ES (2002) Developing passenger car equivalents for heavy vehicles on freeways during queue discharge flow. *Transp Res Part A* 36:725–742
98. Fan S, Work D (2015) A heterogeneous multiclass traffic flow model with creeping. *SIAM J Appl Math* 75:813–835
99. Tuerprasert K, Aswakul C (2010) Multiclass cell transmission model for heterogeneous mobility in general topology of road network. *J Intell Transp Syst* 14:68–82
100. Levin MW, Boyles SD (2016) A multiclass cell transmission model for shared human and autonomous vehicle roads. *Transp Res Part C* 62:103–116
101. Qian Z, Li J, Li X, Zhang M, Wang H (2017) Modeling heterogeneous traffic flow: a pragmatic approach. *Transp Res Part B* 99:183–204
102. Liu H, Wang J, Wijayaratna K, Dixit VV, Travis Waller S (2015) Integrating the bus vehicle class into the cell transmission model. *IEEE Trans Intell Transp Syst* 16:2620–2630