# Contents

1 Introduction ............................................................................................................. 1
   1.1 Review of some traffic flow models ................................................................. 2
      1.1.1 Non-physical queue models .................................................................. 2
      1.1.2 Physical-queue models ........................................................................ 4
   1.2 Examples of traffic modeling ............................................................................ 6
      1.2.1 Shock wave ......................................................................................... 6
      1.2.2 Spillback ............................................................................................. 6
      1.2.3 Gridlock .............................................................................................. 9
      1.2.4 Traffic signal control .......................................................................... 10
      1.2.5 Car trajectories .................................................................................. 14
      1.2.6 Second-order extension of the LWR model ....................................... 15
   1.3 State of the art and challenges ........................................................................ 17
   1.4 The materials ahead ....................................................................................... 19

2 Macroscopic and Mesoscopic Traffic Models .................................................. 21
   2.1 The LWR model ............................................................................................ 22
      2.1.1 Derivation of the equation ................................................................. 22
      2.1.2 Fundamental diagrams ...................................................................... 23
      2.1.3 Riemann problems ............................................................................. 25
      2.1.4 LWR model with viscosity .................................................................. 26
      2.1.5 Hamilton-Jacobi formulation ............................................................. 28
      2.1.6 LWR model in other coordinate systems .......................................... 30
   2.2 The Payne-Whitham model .......................................................................... 33
   2.3 Drawbacks of second order models ............................................................. 35
   2.4 The Aw-Rascle-Zhang model .................................................................... 36
      2.4.1 Characteristic fields .......................................................................... 37
      2.4.2 Invariant domains .............................................................................. 37
   2.5 Third order models ....................................................................................... 39
   2.6 Hyperbolic phase transition model ............................................................. 40
      2.6.1 The Riemann problem ..................................................................... 42
   2.7 The Greenberg-Klar-Rascle multilane model ............................................. 44
### Contents

2.8 A multipopulation model ........................................... 45  
2.8.1 The case \( n = 2 \) ........................................... 47  
2.9 A multiclass model with creeping .................................. 48  
2.10 The Aw-Rascle-Zhang model with phase transition .............. 49  
2.11 The Siebel-Mauser balanced vehicular traffic model .......... 49  
2.12 Generalized Aw-Rascle-Zhang models ............................. 50  
2.13 No-gap phase transition models .................................. 51  
2.13.1 LWR and follow-the-leader model ....................... 53  
2.14 Qualitative features of models for urban traffic ............... 54  
2.14.1 First order models ........................................... 55  
2.14.2 Second order models ......................................... 57  
2.14.3 The Helbing third order model ............................... 63  
2.14.4 Conclusions .................................................. 66  
2.15 Mesoscopic models ................................................ 67  
2.15.1 The Prigogine model ......................................... 68  
2.15.2 The Paveri-Fontana model ................................... 69  
2.15.3 Enskog-like models .......................................... 69  
2.15.4 Discrete velocity models ..................................... 70  
2.16 Bibliographical note ................................................ 71  

3 Models for Traffic on Networks ........................................ 73  
3.1 Discrete models ..................................................... 74  
3.1.1 The cell transmission model .................................. 74  
3.1.2 The link transmission model ................................... 78  
3.2 Single conservation law models ..................................... 80  
3.2.1 Solution for \( 1 \times 1 \) junction ............................ 83  
3.2.2 Solution for \( 2 \times 1 \) (merge) junction .................... 83  
3.2.3 Solution for \( 1 \times 2 \) (diverge) junction .................... 85  
3.2.4 Solution for \( 2 \times 2 \) junction ............................. 86  
3.2.5 A solver redirecting traffic ................................. 87  
3.2.6 A solver keeping priorities and not redirecting traffic .... 89  
3.2.7 A solver for T-junctions ...................................... 90  
3.3 Comparison of discrete and conservation laws models .......... 91  
3.3.1 Comparison of the CTM with the conservation law model .. 92  
3.3.2 Comparison of the LTM with the variational formulation ... 93  
3.4 The Aw-Rascle-Zhang model on networks ......................... 100  
3.4.1 A model conserving the number of cars ..................... 100  
3.4.2 A model conserving the number of cars and the generalized momentum ........................................ 102  
3.5 Phase transition models on networks ............................... 103  
3.6 Coupling of models ................................................ 105  
3.6.1 Follow-the-leader model ..................................... 106  
3.6.2 Follow-the-leader and LWR ................................ 106  
3.6.3 Follow-the-Leader and ARZ ................................. 107
6.4 Continuity of the path delay operator .................................................. 218
6.4.1 The two junction models ................................................................. 219
6.4.2 Well-posedness of the diverge junction model ................................. 221
6.4.3 Well-posedness of the merge junction model ..................................... 226
6.4.4 Continuity of the delay operator ...................................................... 226
6.5 Existence of dynamic user equilibrium ................................................. 235
6.5.1 Preliminaries ................................................................................. 235
6.5.2 The existence result ....................................................................... 236
6.6 Computation of dynamic user equilibrium ............................................ 242
6.6.1 The differential variational inequality formulation ......................... 242
6.6.2 Solution algorithm based on the fixed-point iterations ..................... 243
6.7 Bibliographical note ............................................................................ 246

7 Control of Traffic at Junctions ................................................................. 247
7.1 Modeling signalized intersections ......................................................... 248
7.2 The on-and-off and the continuum signal models ................................. 250
7.2.1 The Hamilton-Jacobi representation of signal models ...................... 251
7.2.2 When spillback is absent ................................................................. 252
7.2.3 When spillback is present and sustained ........................................... 255
7.2.4 When spillback is present and transient .......................................... 264
7.3 Optimization of traffic signal timing ...................................................... 266
7.3.1 Mixed integer linear program approach ......................................... 267
7.3.2 Metaheuristic approach .................................................................. 270
7.4 Flow models for lights vs circles ............................................................ 275
7.4.1 Flux control with traffic lights .......................................................... 275
7.4.2 Single-lane traffic circle with low traffic ......................................... 279
7.4.3 Single-lane traffic circle with heavy traffic ...................................... 281
7.4.4 Multi-lane traffic circle with no interaction ..................................... 284
7.4.5 Traffic light vs traffic circle ............................................................. 285
7.4.6 Conclusions .................................................................................. 288
7.5 Bibliographical note ............................................................................. 288

8 Car Paths, Bottlenecks and Flux Limiters .............................................. 289
8.1 Car paths for conservation laws models .............................................. 290
8.1.1 Wave front tracking for car paths ................................................. 290
8.2 Moving bottlenecks ............................................................................ 296
8.2.1 Solutions by fractional step ............................................................ 298
8.3 Fixed bottlenecks and flux limiters for the LWR model ....................... 302
8.3.1 Flux limiters as 1-1 junctions with flux constraint ......................... 302
8.4 Flux limiters for the Aw-Rascle-Zhang model ...................................... 304
8.4.1 The first constrained Riemann solver .......................................... 306
8.4.2 The second constrained Riemann solver ....................................... 306
8.5 Moving bottlenecks via flux limiters ..................................................... 307
8.6 Bibliographical note ............................................................................ 308
The modeling of dynamic traffic flow on networks has been an important aspect of many scientific inquiries and engineering applications. Traffic flow models based on the conservation of vehicles, from the 1950’s forward, have been greatly influenced by the Lighthill-Whitham-Richards (LWR) model proposed by [199] and [240]. Notably, for the past few decades the LWR model, along with its discrete-time variations, have evolved from ones developed to describe the relationship among aggregated quantities of flow, density and velocity, to new model generations. The latter include models that are network based, provide critical information on speed and acceleration to allow accurate estimation of emissions, treat vehicle spillback and gridlock phenomena, and facilitate engineering applications such as traffic monitoring and control.

The majority of these models focus on the following primary areas:

1. network extension of the fluid-based models [52, 108, 143, 155, 164, 171, 172];
2. numerical approximation and/or discrete-time representation of the traffic dynamics [39, 74, 75, 126, 221, 222, 223, 275];
3. higher-order extension of the LWR scalar conservation law model [10, 24, 54, 55, 60, 76, 115, 278];
4. drivers’ departure time and/or route choices [31, 32, 33, 96, 98, 138, 206, 231, 232, 232].

The fourth item is extensively studied as dynamic traffic assignment (DTA), which is the descriptive modeling of time-varying traffic flows on networks consistent with the notions of travel demand theory and traffic flow theory. The DTA problems take on two primary forms: dynamic user equilibrium (DUE) and dynamic system optimal (SO-DTA); and they correspond respectively to Wardrop’s first and second principles [268]. DUE assumes that travelers are selfish Nash agents, and seeks to find a traffic flow pattern such that the resulting travel costs are equal and minimum within any given origin-destination pair regardless of the route and/or departure time choices. SO-DTA, on the other hand, stipulates that the network-wide total travel cost is minimized,
implying that travelers behave in a cooperative way. The integration of traffic flow theory with travel choice principles, along with various extensions related to demand elasticity, bounded rationality, stochasticity, feedback mechanism, and information and communication technology, form the foundation of DTA, in which materials covered in this book play an underpinning role.

1.1 Review of some traffic flow models

This section offers an overview of some of the most widely used traffic flow models, along with their mathematical expressions, properties, and applications. In presenting these models, we take a common perspective of distinguishing between non-physical queue models and physical queue models, whose meanings will be explained below.

1.1.1 Non-physical queue models

Non-physical queue models typically involve the assumption that, regardless of the severity of the congestion on a link, vehicles on that link do not spill into upstream links. Non-physical queue models assume, either explicitly or implicitly, that links have infinite holding capacities and hence queuing on one link do not affect other links in the network. Examples of such models include the Merchant-Nemhauser model [214, 215], the link delay model [96], and the Vickrey (point queue) model [139, 140, 267].

The Merchant-Nemhauser model

Merchant and Nemhauser proposed a link performance model based on a link exit function, in their seminal work on system-optimal dynamic traffic assignment (DTA) [214]. This model, hereafter referred to as the M-N model, assumes that the flow at which cars leave a link can be expressed as a function of the link occupancy, i.e. the number of cars present on the link. To express this hypothesis symbolically, let $x(t)$ be the number of cars (occupancy) on a given link at time $t$, and $f = f(x)$ be an exit function that represents the flow at which vehicles leave this link. In addition, let $e(t)$ be the rate at which cars enter this link. Then the flow conservation equation can be written as

$$\frac{dx}{dt} = e(t) - f(x(t)).$$

The M-N model also uses a static link performance function to represent the travel cost as a function of link occupancy. The M-N model has inspired several in-depth studies of DTA problems, such as [43, 44, 100].
1.1 Review of some traffic flow models

The link delay model

The link delay model (LDM) was first introduced in [96] as part of a mathematical formulation of dynamic user equilibrium. The LDM is based on a link travel time function $D = D(x)$, that depends on the occupancy of the link. Mathematically, we let $x(t)$ be the occupancy of the link, and define a link travel time function $D(x(t))$. Moreover, let $f(t)$ and $g(t)$ be the entering and exiting flows at time $t$, respectively. Then, an obvious identity is

$$\frac{dx}{dt} = f(t) - g(t).$$

Moreover, the exit time of a car that enters the link at time $t$, denoted by $\tau(t)$, satisfies

$$\tau(t) = t + D(x(t)).$$

A flow propagation constraint for ensuring the First-In-First-Out (FIFO) principle was suggested as

$$F(t) = G(\tau(t)) = G(t + D(x(t)))$$

where $F(t)$ and $G(t)$ represent cumulative numbers of vehicles that have entered and exited the link, respectively. In other words, they satisfy

$$F(t) = \int_0^t f(s) \, ds, \quad G(t) = \int_0^t g(s) \, ds.$$

It has been widely recognized that the FIFO rule is obeyed if the link travel time is an affine function of the link occupancy. However, no other form of the travel time function has been identified for the LDM that is consistent with the FIFO rule ([224, 273]).

The Vickrey model

This model was proposed by Nobel Laureate William S. Vickrey in [267], and is also known as the point queue model. It is widely used in the modeling of traffic networks ([83, 160, 183]) and supply chain networks ([6, 127, 141]). The Vickrey model is based on the assumption that the queue has zero physical size and that the travel time on the link consists of a free-flow travel time plus a congestion related queuing time. A commonly adopted mathematical formulation of the Vickrey model is an ordinary differential equation:

$$\frac{dq(t)}{dt} = f(t) - \begin{cases} \min\{f(t-t_0), C\} & \text{if } q(t) = 0 \\ C & \text{if } q(t) > 0 \end{cases} \quad (1.1.1)$$

where $q(t)$ denotes the size of the queue located at the exit of the link, $t_0$ is the free-flow travel time, $f(t)$ denotes the entering flow, and $C$ is the maximum flow at which the link discharges vehicles (flow capacity). This ODE
has a discontinuous right hand side, which leads to a number of theoretical and numerical difficulties, as discussed by [11]. Notably, Han et al. [139, 140] provide the following closed-form solution of this ODE by reformulating it as a partial differential equation and applying a variational method known as the Lax-Hopf formula:

\[ q(t) = F(t - t_0) - C(t - t_0) - \min_{0 \leq \tau \leq t - t_0} \{ F(\tau) - C\tau \} \]  

(1.1.2)

where

\[ F(t) = \int_0^t f(s) \, ds \]

1.1.2 Physical-queue models

Another important class of traffic flow models, the physical-queue models, are able to capture the formation, propagation and dissipation of queues and allow spillback to be explicitly represented. Such models are arguably more realistic than non-physical queue models. Among such physical queue models are the LWR model, and two of its discrete-time variations: the cell transmission model [74, 75] and the link transmission model [275].

The LWR model

The LWR model describes the spatial-temporal evolution of vehicle density on a link using a scalar conservation law:

\[ \partial_t \rho + \partial_x f(\rho) = 0 \]  

(1.1.3)

where the flux function \( f(\cdot) \) is assumed to be continuous, concave, and vanishes at \( \rho = 0 \) and \( \rho = \rho^{jam} \); the latter is the jam or maximal density. The graph of the function \( \rho \mapsto f(\rho) \) is called the fundamental diagram. A few most widely used fundamental diagrams are the Greenshields fundamental diagram [132], the trapezoidal fundamental diagram [74, 75], and the triangular fundamental diagram [221].

The LWR model is derived based on the conservation of cars and an empirical relationship between vehicle density and flow, which is expressed by the fundamental diagram. The LWR model captures some realistic traffic phenomena such as shock waves, physical queues, and spillback. More detailed introduction of this model, along with its mathematical properties and network extensions, will be presented in later sections of this book.

The cell transmission model

The cell transmission model (CTM), proposed by the two seminal papers of Daganzo [74, 75], presents a simplified and discrete-time version of the
1.1 Review of some traffic flow models

LWR model. The CTM relies on a trapezoidal or triangular fundamental diagram, and is shown to be consistent with the Godunov discretization scheme \([126, 190]\) for solving scalar conservation laws of the form \((1.1.3)\). It not only reduces the partial differential equation model to a set of simple procedures that facilitate computation through straightforward bookkeeping, but also allows the model to be generalized to a traffic network through the notion of demand and supply functions; see \([74, 75, 191]\).

The model works as follows. The road segment of interest is first partitioned into a sequence of spatial intervals, called cells, labeled by the index \(j\). We use the index \(i\) to denote discrete time steps (intervals). The propagation of traffic in the space-time diagram is governed by the following set of equations

\[
\begin{align*}
    n_j(i + 1) &= n_j(i) + y_j(i) - y_{j+1}(i) \\
    y_j(i) &= \min \left\{ n_{j-1}(i), Q_j(i), \frac{w}{v}(N_j(i) - n_j(i)) \right\}.
\end{align*}
\]

Here \(n_j(i)\) denotes the number of vehicles in cell \(j\) during time interval \(i\); \(y_j(i)\) denotes the number of vehicles entering cell \(j\) from its upstream neighboring cell \(j - 1\) during time interval \(i\); \(Q_j(i)\) is the maximum number of vehicles that can be discharged into cell \(j\) from cell \(j - 1\), during the time interval \(i\); \(N_j(i)\) is the holding capacity of cell \(j\) at time \(i\). The positive constants \(w\) and \(v\) are the forward and backward wave speeds, respectively, that are associated with the triangular or trapezoidal fundamental diagram. A more elaborated discussion of this model and its network extension will be presented later in Section 3.1.1.

The CTM is known to coincide with the Godunov discretization scheme of the LWR conservation law, and this will be illustrated in Section 3.3.1.

The link transmission model

The link transmission model (LTM) was originally proposed in \([276]\). It is an efficient implementation of Newell’s variational theory for kinematic wave models \((221, 222, 223)\) on a traffic network. Unlike the CTM, which keeps track of the occupancy of each cell, the LTM consider quantities that are associated with an entire link and stipulates rules for sending and receiving flows between links. As a result, the computational efforts related to memory usage and numerical operations are substantially reduced by the LTM. Notably, the LTM also admits a network extension.

The variational theory employed by the LTM is based on the following Hamilton-Jacobi equation, closely related to the conservation law \((1.1.3)\):

\[
\partial_t N(t, x) - f(-\partial_x N(t, x)) = 0
\]

where \(N(t, x)\), called the Moskowitz function \([217]\), measures the cumulative number of vehicles that have passed location \(x\) along a road segment by time \(t\). The variational approach for solving the above Hamilton-Jacobi equation
has been derived independently from calculus of variations ([87]), the viability theory ([8]), and traffic flow theory ([77, 221]). The variational theory gives rise to an analytic representation of the Moskowitz function, which is equivalent to a minimization problem. If the fundamental diagram is triangular or piecewise affine, then the solution of such a minimization problem can be expressed in closed form. We will present this model in further mathematical detail in Sections 3.1.2 and 3.3.2.

1.2 Examples of traffic modeling

A congested traffic network often exhibits complex behavior induced by physical constraints (e.g. road capacity, speed limit), active control measures (e.g. signal control, ramp metering), and diverse characteristics of drivers (e.g. driving behavior, origin-destination). Modern traffic flow models, particularly conservation law models, are capable of capturing one or several realistic phenomena commonly seen on traffic networks using mathematical tools. In this section, we will provide several examples of these phenomena and illustrate their mathematical models.

1.2.1 Shock wave

Shock wave refers to a (possibly) moving boundary that separates two distinct vehicle density values. It is described as a sharp transition from one density state to another, e.g. when a car travels with free-flow condition before joining a queue. The shock wave is often represented mathematically as a jump discontinuity of car density in space, which is a natural consequence and a defining characteristic of conservation law models. Various examples of shock waves are provided in the remainder of this chapter (e.g. Sections 1.2.2, 1.2.4 and 1.2.5) and throughout this book.

1.2.2 Spillback

Vehicle spillback is a phenomenon commonly seen in congested road networks. It typically happens when the entrance of a link becomes congested, causing its receiving capacity to drop and affecting the upstream road junction as well as its immediately preceding links. It has been well understood, both empirically and numerically, that spillback is a major source of network-wide congestion and gridlock, and that it causes additional travel delays that are not captured by non-physical queue models.

We use Figure 1.1 below to illustrate spillback and its impact on path travel times. The picture on the left hand side, obtained using the LWR model, depicts a diverge junction where link $c$ is heavily congested and link $b$ is empty. Moreover, assume that all vehicles on link $a$ intend to turn right into link $c$.
1.2 Examples of traffic modeling

except one vehicle $x$, which is represented as the red dot. The congestion on link $c$ causes its receiving capacity to drop, restricting the cars on link $a$ from exiting the link. As a result, a queue is formed on link $a$ and prevents the car $x$ from entering the empty link $b$.

![LWR model and Point queue model](image)

**Fig. 1.1.** Example of spillback. Evolution captured by the physical-queue model (left) and not captured by the point-queue model (right).

We now turn to the right hand side of Figure 1.1 which shows a similar situation but assuming the point queue (Vickrey) model. In this case, the congestion on link $c$ is concentrated on the point (non-spatial) queue, leaving the junction clear for car $x$ to enter link $b$.

A comparison between the two cases in Figure 1.1 leads to the following conclusion:

1. spillback is responsible for the propagation among different links in the network (e.g., from link $c$ to $a$);
2. the presence of spillback leads to very different path travel times compared to the non-physical queue models.

For this reason, spillback is considered one important feature of a realistic dynamic traffic network model. In addition, it has a few other consequences such as gridlock, which will be illustrated in the next subsection.

In order to show spillback in a more visualizable way, we compute a solution of the LWR model on the following network presented in Figure 1.2. Details of the model are postponed in Chapter 3.

In Figure 1.3 we show the space-time diagram for the car density on links $I_3$ and $I_1$. We can see that link $I_3$ is clearly separated into two regions by a shock wave; the one to the left is the free-flow region and the one to the right is the congested region. The shock wave originates from the downstream boundary of the link and propagate at varying speed into the link. At around $t = 3.5$ (hour) the shock reaches the upstream boundary of $I_3$ and this is when spillback occurs at node 2 (see Figure 1.2). We can see that a shock wave separating the free-flow region and the congested region emerges at the downstream boundary of link $I_1$, as if the congestion ‘penetrates’ the mutual
boundary of $I_3$ and $I_1$. This is a clear visualization of the propagation of congestion between different links in the network. Similarly, Figure 1.4 shows spillback from link $I_4$ into link $I_2$.

**Fig. 1.2.** The seven-arc test network.

**Fig. 1.3.** Illustration of spillback from link $I_3$ into $I_1$.

**Fig. 1.4.** Illustration of spillback from link $I_4$ into $I_2$.
1.2.3 Gridlock

Another possible consequence of the physical-queue models is network gridlock. Gridlock typically arises when there exists a circle in the network and, as the level of saturation grows, vehicle spillback forms a loop. As a result, the throughput of this local subnetwork keep decreasing to the point where cars can hardly move and the car densities on relevant links are maximum.

Gridlock is a realistic traffic phenomenon, which has been observed empirically and captured by conservation law models. We use the LWR model to illustrate the gridlock phenomenon on the traffic roundabout shown in Figure 1.5 the reader is referred to Section 6.3 for details of the model. In this network, vehicles are associated with three origin-destination pairs: (1, 6), (3, 2), and (5, 4), and they follow the unique, non-circular path from the origin to the destination (i.e. traversing two-thirds of the circle).

The network is initially empty. We set the departure rates between all three origin-destination pairs to be the same constant: 800 vehicle/hour (all links have the same flow capacity of 3000 vehicle/hour). It is expected that congestion will originate from the downstream end of each link on the circle and propagate in the counter-clockwise direction. Then, spillback will occur simultaneously at the three nodes on the circle. As congestion accumulates, the car density on the circle approaches the jam density and the throughput of the network becomes smaller. Figure 1.6 shows the flow at which vehicles arrive at the destination 6 (the situations for the other destinations are identical).

The gridlock effect can be clearly seen from Figure 1.6 where the outflows of all three links on the circle tend to zero as time advances, and the density on each link approaches a bumper-to-bumper jam state. And this is taking place even though the network overall is under saturated: The departure rates of 800 vehicle/hour are well below the link capacity of 3000 vehicle/hour.
1.2.4 Traffic signal control

Signalized intersections play a vital role in the design, management, and control of urban traffic networks. They not only have a direct impact on congestion and journey time, but are also emission hotspots due to vehicle acceleration/deceleration concentrated near the stop line. Therefore, urban traffic modeling and control tend to focus on the operation of signalized links and intersections.
The modeling of a signalized road link is conceptually simple: vehicles are allowed to move through the signal during the green phase or ‘on’ period and are restricted from moving during the red phase or ‘off’ period; and the periodic switching between these two distinct phases must be tracked to accurately model vehicle throughput. We consider a link whose dynamic is described by the LWR model with a Greenshields (quadratic) fundamental diagram (132):

$$f(\rho) = \rho \cdot \left(1 - \frac{\rho}{\rho_{jam}}\right) \cdot v_0$$

where \(\rho\) denotes vehicle density, \(\rho_{jam}\) denotes the maximum (jam) density, and \(v_0\) is the free-flow speed. The Greenshields fundamental diagram is shown in Figure 1.8.

Assume that a periodic signal control is acting on this link, and that the downstream of the link has sufficient capacity so that cars can be discharged at the maximum rate \(C\) when the green light is on. The kinematic wave profile on this link is illustrated in the space-time diagram in Figure 1.9.

In Figure 1.9, the flow into the subject link is assumed to be a constant \(q^{in}\), and is represented as the \(\rho^{in}\)-wave near the entrance of the link, where

$$\rho^{in} \leq \rho_{jam}/2, \quad f(\rho^{in}) = q^{in}.$$
Near the exit of the link, the vehicle density profile is influenced by the on-and-off signal sequence. More specifically, when the light is red the discharge flow is zero, corresponding to a jam density, which propagates backward with speed $v_0$. This $\rho_{jam}$-wave interacts with the constant density $\rho^{in}$, creating a shock wave with speed given by the Rankine-Hugoniot condition:

$$\frac{f(\rho_{jam}) - f(\rho^{in})}{\rho_{jam} - \rho^{in}} = \frac{-q^{in}}{\rho_{jam} - \rho^{in}} < 0.$$
When the light turns green, vehicles are discharged with the saturation flow rate $C$, corresponding to the critical density $\rho_{jam}^2$, which propagates with zero speed. The area between this zero-speed wave and the $\rho_{jam}$-wave from the previous red phase is filled with rarefaction waves (expansion wave or fan wave). These rarefaction waves create two more shock waves, with $\rho^{in}$ and $\rho_{jam}$ respectively. These two shock waves have varying speeds since the density values corresponding to the rarefaction waves are changing; and one has to solve an ordinary differential equation (ODE) to obtain a closed-form representation of the shock speeds.

Figure 1.10 shows the numerically computed density profile on the link, where $\rho_{jam} = 400$ (vehicle/km), $v_0 = 48$ (km/hour), and $q^{in} = 2400$ (vehicle/hour). The initial density on this link is uniformly 200 (vehicle/km) and the signal control starts with the red phase. In this figure we can clearly observe shock waves represented as sharp transitions from one density to another, and rarefaction waves represented as smoothly varying densities.

The entire link is separated by a shock wave into the uncongested region (with density $\rho^{in}$) and the congested region. This shock wave moves back and forth as time advances, indicating the recurring growth and dissipation of queues. Moreover, one could show, through nontrivial calculation involving ODEs, that this shock wave remains away from both the entrance and exit of the link if and only if

$$q^{in} = f(\rho^{in}) = \frac{\Delta_g}{\Delta} \cdot C$$

where $\Delta$ and $\Delta_g$ denote the cycle length and the green time per cycle of the signal control, respectively. And, the shock will move towards the entrance.
1 Introduction

(exit) of the link if \( q^{in} > \frac{A}{2}c \) (\( q^{in} < \frac{A}{2}c \)). These results may be derived using a different approach, by considering the average throughput of the link given by \( \frac{A}{2}c \). Thus the congestion of the link will grow (or dissipate) if the link inflow \( q^{in} \) is greater (or less) than \( \frac{A}{2}c \).

The latter perspective is based on a temporal aggregation of the signal control, in which the on-and-off (binary) signal control is replaced by its weak limit \( \frac{A}{2}c \in (0, 1) \) when the cycle length tends to zero or the time horizon tends to infinity. Such temporal homogenization gives rise to a new class of signal control models known as the continuum signal models \((143, 142)\). More details of this model, including its theoretical and computational issues, will be presented in Chapter 7.

1.2.5 Car trajectories

With the increased availability of mobile traffic data and the advancement of sensing technology, data collected through GPS, smartphones or other mobile devices have become a major source of traffic information for various applications. Modern macroscopic traffic models are challenged by the need to accurately capture and predict individual vehicle trajectories on the link, corridor, and network levels. Such information is not immediately available from the LWR type models as car speeds are different from kinematic wave speeds (the former are greater than or equal to the latter).

One way to infer individual car trajectories from the LWR type models is to invoke the Moskowitz function \((217)\), also known as the Newell curve \((221, 222, 223)\), \( N(t, x) \), which represents the cumulative number of vehicles that have passed certain location \( x \) along the link by time \( t \). By definition \( N(t, x) \) is monotonically increasing in \( t \) and monotonically decreasing in \( x \). In fact, the following identities hold:

\[
\partial_t N(t, x) = f(\rho(t, x)), \quad \partial_x N(t, x) = -\rho(t, x)
\]

where \( \rho(t, x) \) represents density and \( f(\rho(t, x)) \) represents flow. The Moskowitz function is closely related to the conservation law as it satisfies a Hamilton-Jacobi equation and can be expressed semi-analytically using the variational principle (Lax-Hopf) formula; more details will be presented in Section 6.3.2.

The function \( N(t, x) \) is naturally related to car trajectories as its graph can be seen as a two-dimensional surface in a three-dimensional Euclidean space. In particular, its contour line \( \{(t, x) : N(t, x) = n\} \), which is parameterized by the Lagrangian label \( n \), represents the time-space trajectory of the moving car labeled as \( n \).

Figure 1.11 shows the contour lines of the Moskowitz function for the same scenario depicted in Figure 1.10. The total number of cars involved in this space-time diagram is 500, and they are labeled as 1-500. Cars that enter the link first have a smaller label; thus the number \( n \) increases from the upper-left
corner to the lower-right corner. In this figure, area $A$ indicates vehicle movements at a constant speed because the density is uniform; area $B$ corresponds to the jam density and zero vehicle speeds; area $C$ represents the rarefaction wave, in which cars gradually accelerate as a result of queue dissipation. The black solid lines represent shock waves, which indicate a sharp change from lower density to higher density in the direction of travel. As a result a kink in the car trajectory can be seen across the shock waves corresponding to the breaking maneuver. It can be also seen from these contour lines that a car typically experiences 2-3 full stops on this link.

![Contour Lines](image)

**Fig. 1.11.** The contour lines of the Moskowitz function $N(t, x)$, which represent car trajectories.

The construction of Figure 1.11 involves solving the Hamilton-Jacobi equation for $N(t, x)$ (see Section B.3 of this book for more detail), and computing contour lines. These operations are in general time consuming and largely affected by the spatio-temporal resolution of the model. A computationally more practical procedure for calculating car trajectories, especially through a road network, will be derived later in this book, in Section 8.1.

### 1.2.6 Second-order extension of the LWR model

Despite the popularity of the LWR scalar conservation law model due to its simple mathematical representation and the capability to capture shock waves, queue spillback and so forth, it has a few limitations, especially in describing and predicting complex traffic waves/phenomena observed in traffic. These include stop-and-go waves, phantom jam, and capacity drop; see [23, 187]. These phenomena are mainly caused by the instability of congested traffic...
and heterogeneous driving behavior, which are insufficiently captured by the single-valued fundamental diagram stipulated by the LWR model.

Driven by the need to capture more realistic traffic behavior especially in long and congested highways, second-order or higher-order conservation law models have been proposed since the 1970s, beginning with the Payne-Whitham model \cite{229, 269}. In a celebrated paper \cite{76} Daganzo criticized second-order models by showing various drawbacks including the possibility of cars going backward. Most of such drawbacks were later addressed by the Aw-Rascle-Zhang model, independently proposed by \cite{10} and \cite{278}. More recently, the phase transition models \cite{54, 60, 115} drew increased attention from researchers for their capability of representing complex waves while keeping the LWR structure for light traffic. Most second-order traffic models tend to pose, in addition to the LWR-type equation for the conservation of vehicles, a second equation for the conservation or balance of momentum.

Let us motivate second-order extension of the LWR model with Figure 1.12, which shows empirical data on the density-flow and density-velocity relationships collected on a segment of the northbound of I-80 located in Emeryville, CA. The dataset mainly focuses on congested traffic, which is shown as the cluster of points for density within \([0, 0.2]\) (in vehicle/foot). We can see that instead of forming a clear one-to-one relationship between density and flow (or velocity) as predicted by the LWR model, the data points are scattered in a two-dimensional region. In other words, the same density value may correspond to a range of values for the speed and flow; as a result, traffic may deviate from the equilibrium state that is captured by the LWR model.

![Figure 1.12](image_url)

**Fig. 1.12.** The density-flow (left) and density-velocity (right) data plots.

The second-order phase transition model uses a 2-D region to describe the fundamental diagram, thus capturing the deviation of traffic from the equilibrium state. After applying certain techniques proposed in \cite{234}, one may calculate the level of traffic deviation for this congested highway segment, which is shown in Figure 1.13. The top picture shows vehicle density during a 30-minute period while the bottom picture shows the deviation from
equilibrium state, which can be both positive and negative and is normalized to be within $[-1, 1]$. Positive (negative) deviation means that the actual speed is higher (lower) than the equilibrium speed. Indeed, we can see that significant deviation occurs where traffic is congested and unstable, and both above-equilibrium and below-equilibrium speeds are present.

The modeling, prediction, and management of highway traffic often require traffic models to capture the aforementioned complex phenomena while remain theoretically and computationally tractable. Some of these issues will be addressed by this book in Chapter 2.

**Fig. 1.13.** Top: the space-time diagram for the density. Bottom: the space-time diagram for the normalized deviation from the equilibrium traffic state.

The modeling, prediction, and management of highway traffic often require traffic models to capture the aforementioned complex phenomena while remain theoretically and computationally tractable. Some of these issues will be addressed by this book in Chapter 2.

1.3 State of the art and challenges

The increasingly important role that conservation law models play in the modeling and simulation of traffic networks comes with challenges and obstacles
Introduction

that need to be properly addressed. Some of the challenges arise from the theoretical aspect of traffic models such as existence, uniqueness, and well-posedness of PDE models; others are driven by practical applications, e.g., integration of conservation law models with massive and heterogeneous traffic data, and real-time and robust decision making for the control and management of large-scale urban networks.

In the following discussion, we list some key challenges in the modeling of traffic networks through conservation law models that are currently faced by the mathematical and engineering research communities. These challenges have been (partially) addressed by this book; and we provide references to relevant chapters/sections after the statement of each problem.

- How to use conservation law models, which are essentially macroscopic, to properly capture vehicle dynamics and interactions that occur on a microscopic level, such as different vehicle types, heterogeneous driving behavior, and emission and fuel consumption rates? These issues are addressed in Chapter 2 where higher-order conservation law models are discussed.
- What are the most appropriate models for traffic intersections that incorporate certain junction controls (e.g., signal light, stop sign, right-of-way) while maintaining sound mathematical properties (e.g., existence, uniqueness, consistency, and well-posedness)? These questions are addressed in Chapter 3 and Chapter 7.
- When the conservation law models are extended to road networks, what is the set of minimum requirements that such extensions must satisfy to ensure the physical realism and desirable mathematical behavior of the resulting network models? These challenges are addressed in Chapter 4.
- How to integrate conservation law models with drivers’ route and departure time choices in order to form a descriptive modeling of dynamic traffic flows on networks consistent with the notions of travel demand and travel choice principles? This line of research, known as dynamic traffic assignment (DTA), will be discussed in depth in Chapter 5 and Chapter 6.
- In the DTA literature, it is now understood that the path travel times may depend in a discontinuous way on the path departure rates; this is especially the case when spillback occurs in the network. This fact has posed a fundamental obstacle to the analysis of DTA models. Thus, a logical and important question to raise is whether there exist certain characteristics of a network model for which the aforementioned dependence is continuous, even in the presence of spillback. This question will be carefully analyzed and partially answered in Section 6.4.
- How to develop models with advanced computing capabilities to support efficient and robust decision making for real-time and large-scale traffic control and management. This challenge is partially addressed in Chapter 7 and Chapter 10 in the context of traffic signal control.
- How to fuse potentially large, multi-source, and heterogeneous traffic datasets into conservation law models and their network extensions? What is the
1.4 The materials ahead

The rest of this book is organized as follows.

Chapter 2 focuses on first- and higher-order macroscopic and mesoscopic conservation law models on links or highway segments, including their origins, main features, mathematical representations, drawbacks, and comparisons.

Chapter 3 presents a theory of traffic network modeling based on the articulation of link dynamics and junction models. It covers a wide spectrum of continuous- and discrete-time; macroscopic and microscopic; and first- and high-order models.

Chapter 4 is concerned with the theoretical aspects of network traffic models. Rigorous results regarding solutions existence, uniqueness, and well-posedness are established based on the mathematical tools of Riemann problems, wave front tracking algorithms, and generalized tangent vectors, etc.

Chapter 5 extends the traffic network models to include sources and destinations of traffic. This is the first chapter in this book to distinguish traffic by their trip purposes, and sets the foundation for the study of dynamic traffic assignment presented later.

Chapter 6 introduces an active field of inquiry in the engineering research community known as dynamic traffic assignment. It builds on the traffic network models presented in this book so far and incorporates drivers’ route and departure time choice principles. This chapter primarily focuses on the theory and computation of dynamic user equilibrium.

Chapter 7 discusses several examples of traffic control and regulation at junctions, including traffic signal modeling and optimization, as well as roundabout design and management.

Chapter 8 considers several scenarios involving heterogeneous network environments such as moving bottlenecks, car paths, and flux limiters. Modeling issues arising therein are discussed with analyses and solutions provided.

Chapter 9 focuses on the numerical implementation of computational algorithms for traffic flow on networks. We demonstrate the performance of
various numerical schemes in terms of accuracy, convergence, and efficiency.

**Chapter 10** demonstrates the real-world relevance of the models presented in this book by showing four specific application problems in transportation. These include traffic monitoring and estimation, network signal optimization incorporating drivers’ routing strategies, real-time signal control on networks, and traffic data fusion.

**Appendix A** presents the fundamental theory of conservation laws not necessarily within the traffic context. This part serves as supplementary materials to supply essential mathematical details for the rest of the book.

**Appendix B** is mainly concerned with the Hamilton-Jacobi equation representation of certain conservation laws, which has received much attention in the engineering community. In particular, the variational theory is presented in detail here and has been referenced throughout the rest of the book.