7
Traffic flow theory and modelling

Victor L. Knoop and Serge Hoogendoorn

7.1   INTRODUCTION

When do traffic jams emerge? Can we predict, given certain demand levels, when queuing will occur, how long the queues will be, how they will propagate in space and time, and how long it takes for the congestion to resolve? Why does an overloaded traffic network underperform? This chapter gives a basic introduction in traffic flow theory which can help to answer these kinds of questions.

We start this chapter with explaining how this chapter connects with the other chapters in this book (see Figure 7.1). Top left in the figure the reader recognizes the conceptual model introduced in Chapter 2 in highly simplified form to explain transport and traffic volumes.

One of the results of the interplay between people's and shippers' needs and desires, the locations of activities, and the transport resistance factors (Figure 7.1, top left) is a certain volume of road traffic (Figure 7.1, middle left). Road traffic, and here starts this chapter, can be described using flow variables such as speed and density (Figure 7.1, middle right). The density of traffic is the number of vehicles that is present on a roadway per unit distance. Road traffic flows on certain road stretches during certain time periods can either be free or congested and/ or the flows can be unreliable. In the latter two cases, the transport resistance on these road stretches will be relatively high as explained in Chapter 6 using, among others, concepts such as the value of time and value of reliability. Consequently, high transport resistance implies negative repercussions on road traffic volumes (see the arrow from flow variables to transport resistance, Figure 7.1, top).

To be clear, this chapter focuses solely on the road traffic flow variables and the interactions with aspects as driving behaviour, weather, information technology, and so forth (the grey areas in Figure 7.1). Thus, traffic flow operations on a road facility are explained for a given traffic demand profile. Factors such as weather and information technology (e.g., navigation systems) can influence traffic flow characteristics via driving behaviour. Additionally, policies such as road expansions and traffic management measures can also have an impact on traffic flow operations, either directly or indirectly by influencing driving behaviour.

Traffic flow theory entails the knowledge of the fundamental characteristics of traffic flows and the associated analytical methods. Examples of such characteristics are road capacities, the
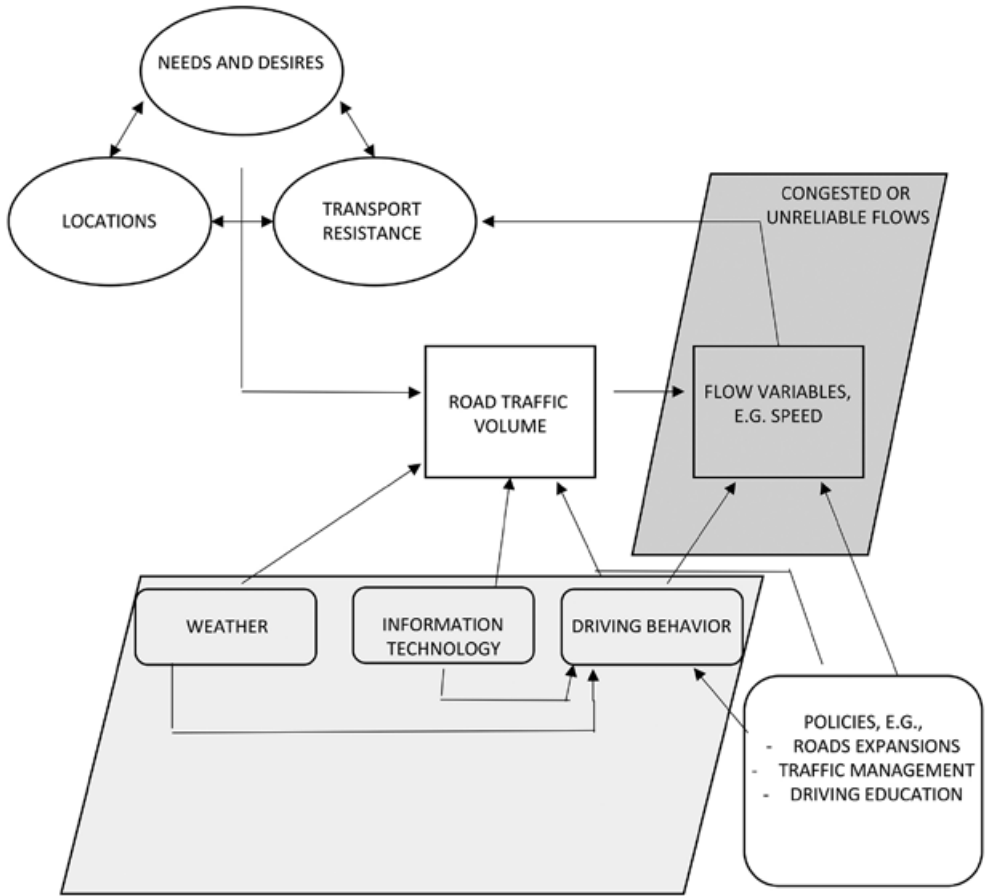
Figure 7.1   The connection of this chapter (grey area) with the simplified conceptual
framework (top left) as described in Chapter 2

relation between flow and density, and headway distributions. Examples of analytical methods
are shockwave theory and microscopic simulation models.

Using the presented material, the reader will be able to interpret, analyse, and – for simple
situations – predict the main characteristics of traffic flows. For the larger part, the chapter
considers traffic flow operations on simple infrastructure elements (uninterrupted traffic
flow operations, simple discontinuities), although an important side step is made to network
dynamics. In doing so, the chapter takes both a *microscopic* and a *macroscopic* perspective.
The microscopic perspective reflects the behaviour of individual drivers interacting with
surrounding vehicles, while the macroscopic perspective considers the average state of traffic.
We discuss empirical facts, and some well known analytical tools, such as shockwave theory,
kinematic wave models, and microscopic simulation models. We will also discuss the appli-
cation of traffic flow theory to bicycle traffic and will consider the impact of the technology of
automated vehicles on the traffic stream.

Section 7.2 introduces the basic variables on the microscopic level (the vehicle level), and Section 7.3 the macroscopic variables (i.e. the flow level). Section 7.4 discusses flow characteristics. Section 7.5 looks into the future developments and discusses the effect of autonomous vehicles. Then, in Section 7.6 traffic flow dynamics and the (self-) organization of traffic are discussed. Section 7.7 presents several theories on multi-lane vehicular traffic (i.e., motorways). Section 7.8 discusses microscopic flow models, while Section 7.9 discusses the macroscopic flow models. Section 7.10 adds the dynamics of networks to this. Section 7.11 shows how all theories and methods can be applied to bicycle traffic. Finally, in Section 7.12 the conclusions are presented.

7.2 VEHICLE TRAJECTORIES AND MICROSCOPIC FLOW VARIABLES

The vehicle trajectory (often denoted as $x_i(t)$) of vehicle (i) describes the position of the vehicle over time (t) along the roadway. The trajectory is the core variable in traffic flow theory which allows us to determine all relevant microscopic and macroscopic traffic flow quantities. Note that for the sake of simplicity, the lateral component of the trajectory is not considered here.

To illustrate the versatility of trajectories, Figure 7.2 below shows several vehicle trajectories. From the figure, it is easy to determine the distance headway S_i, and the time headway h_i, overtaking events (crossing trajectories), the speed $v_i = dx_i/dt$, the size of the acceleration (see top left where one vehicle accelerates to overtake another vehicle), the travel time TT_i, and so forth.

However, although the situation is rapidly changing due to so-called floating car data becoming more common, trajectory information is seldom available. Floating car data is information from mobile phones in vehicles that are being driven. In most cases, vehicle trajectory measurements only contain information about average characteristics of the traffic flow, provide only local information, or aggregate information in some other way (e.g., travel times from automatic vehicle identification or licence plate cameras).

Most commonly, traffic is measured by (inductive) loops measuring local (or *time–mean*) traffic flow quantities, such as (local) traffic flow q, and local mean speed u. First, we will discuss the main microscopic traffic flow variables in detail. This type of flow variables reflects the behaviour of individual drivers interacting with surrounding vehicles.
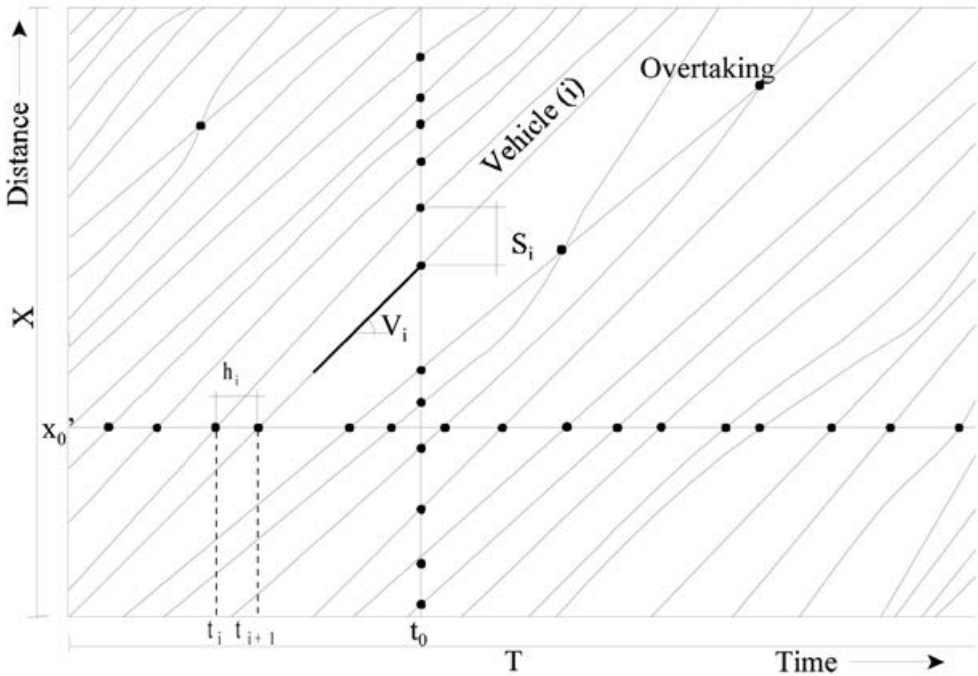
Figure 7.2   Vehicle trajectories and key microscopic flow characteristics

7.2.1   Gross and Net Headways

The (gross) time headway (h) is one of the most important microscopic flow variables. It describes the difference between passage times t_i at a cross-section x of the rear bumpers of two successive vehicles:

$$h_i(x) \; = \; t_i(x) - t_{i-1}(x) \tag{7.1}$$

The time headway, or simply headway, is directly determined by the behaviour of the driver, vehicle characteristics, flow conditions, etc. Its importance stems from the fact that the (minimal) headways directly determine the capacity of a road, a roundabout, and so forth. Typically, these minimal headways are around 1.5 seconds in dry conditions. Time headways, combined with the speeds, leads to the distance headways (see below).

The *net time headway* or *gap* is defined by the difference in passage times between the rear bumper of the lead vehicle and the front bumper of the following vehicle. This value is in particular important for driving behaviour analysis, for instance when analysing and modelling the amount of space drivers need to perform an overtaking manoeuvre (so-called *critical gap analysis*).

7.2.2   Gross and Net Distance Headways

We have seen in the preceding subsection that time headways are *local* microscopic variables: they relate to the behaviour of an individual driver and are measured at a cross-section. On the contrary, distance headways (often denoted with the symbol *s*) are *instantaneous* (measured at one moment in time) microscopic variables, measuring the distance between the rear bumper of the leader and the rear bumper of the follower at time instant *t*:

$$s_i(t) \; = \; x_{i-1}(t) - x_i(t) \tag{7.2}$$

In congested conditions, distance headways are determined by the behaviour of drivers, which in turn depends on the traffic conditions, driver abilities, the vehicle characteristics, weather conditions, and so forth. In free flow with no interaction between the drivers, the headways are determined largely by the demand (that is: they are determined by the moments drivers enter the freeway).

Net distance headways are, similar to the net time headways, defined as the distance between the position of the rear bumper of the leader and the front bumper of the follower.

It should be clear that the time headways and the distance headways are strongly correlated. If v_{i-1} denotes the speed of the leading vehicle, it is easy to see that:

$$s_i \; = \; v_{i-1} h_i \tag{7.3}$$

7.3   MACROSCOPIC FLOW VARIABLES

So far, we have mainly looked at microscopic traffic flow variables. Macroscopic flow variables, such as flow, density, speed, and speed variance, reflect the average state of the traffic flow in contrast to the microscopic traffic flow variables which focus on individual drivers. Let us take a closer look at the most important variables.

7.3.1   Traditional Definitions of Flow, Density, and Speed

In general, the flow *q* (also referred to as intensity, or volume) is traditionally defined by the "average number of vehicles (*n*) that passes a cross-section during a unit of time (*T*)". According to this definition, flow is a *local variable* (since it is defined at a cross-section). We have:

$$q \; = \; \frac{n}{T} \; = \; \frac{n}{\sum_{i=1}^{n} h_i} \; = \; \frac{1}{h} \tag{7.4}$$

This expression shows that the flow can be computed easily by taking the number of vehicles *n* that has passed the measurement location during a period of length *T*. The expression also

shows how the flow q relates to the average headway $\bar{h}$, thereby relating the macroscopic flow variable to average microscopic behaviour (i.e., time headways).

In a similar way, the density k (or *concentration*) is defined by the "number of vehicles per distance unit". Density is, therefore, a so-called *instantaneous variable* (i.e., it is computed at a time instant), defined as follows:

$$k \; = \; \frac{m}{X} \; = \; \frac{m}{\sum_{i=1}^{m} s_i} \; = \; \frac{1}{s} \qquad\qquad (7.5)$$

This expression shows that the density can be computed by taking a snapshot of a roadway segment of length X and counting the number of vehicles m that occupy the road at that time instant. The expression also shows how density relates to average microscopic behaviour (i.e., distance headways, s). Note that contrary to the flow, which can generally be easily determined in practice using cross-sectional measurement equipment (such as inductive loops), the density is not so easily determined since it requires observations of the entire road at a time instant (e.g., via an aerial photograph).

Similarly to the definitions above, average speeds u can be computed in two ways: at a cross-section (local mean speed or time–mean–speed u_L), or at a time instant (instantaneous mean speed or space–mean-speed u_M). As will be shown in the following paragraph, the difference between these definitions can be very large. Surprisingly, in practice the difference is seldom made. For instance, the Dutch motorway monitoring systems collects time–mean speeds, while for most applications (e.g., average travel time) the space–mean speeds are more suitable.
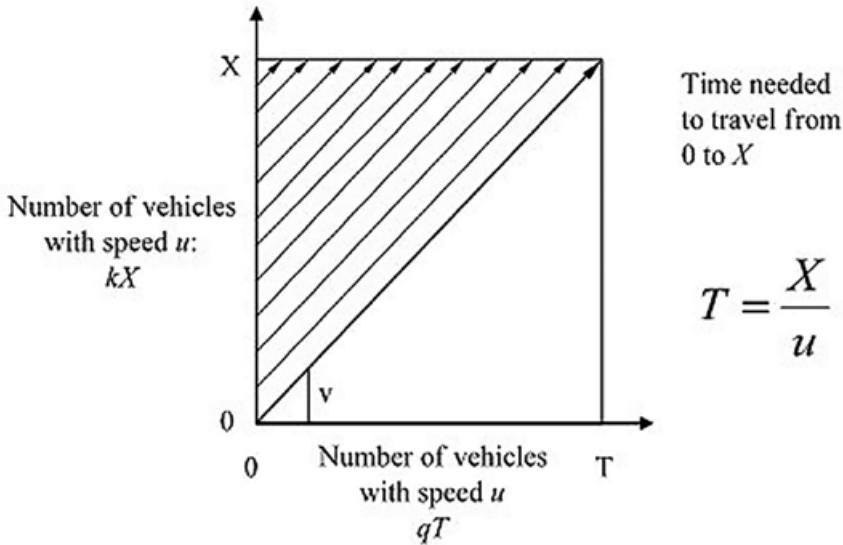
7.3.2   Continuity Equation

An important relation in traffic flow theory is the continuity equation: $q = ku$ (flow equals density times the speed). This equation is used to relate the instantaneous characteristic density to the local characteristic flow. The derivation of this equation is actually quite straightforward (Figure 7.3).

Consider a road of length X. All vehicles on this road drive at an equal speed u. Let us define the period T by $T = X/u$. Under this assumption, it is easy to see that the number of vehicles that are on the road at time $t = 0$ – which is equal to the density k times the length X of the roadway segment – is equal to the number of vehicles that will pass the exit at $x = X$ during period $[0,T]$ – which is in turn equal to the flow q times the duration of the period T. That is:

$$kX \; = \; qT \iff q \; = \; k\frac{X}{T} \; = \; ku \qquad\qquad (7.6)$$

Clearly, the continuity equation holds when the speeds are constant. The question is whether the equation $q = ku$ can also be applied when the speeds are not constant (e.g., u represents an average speed), and if so, which average speed (time–mean or space–mean speed) is to be used. It turns out that $q = ku$ can indeed be applied, but only if $u = u_M$, that is, if we take the space mean speed.

Figure 7.3   Derivation of the continuity equation

Intuitively, one can understand this as follows (a mathematical proof can be found in May, 1990). A detector lies at location x_{det}. Now we reconstruct which vehicles will pass in the time of one aggregation period. For this, the vehicle must be closer to the detector than the distance it travels in the aggregation time t_{agg}:

$$x_{\text{det}} - x_j \leq t_{\text{agg}} v_i \tag{7.7}$$

In this formula, x is the position on the road. For faster vehicles, this distance is larger. Therefore, if one takes the local arithmetic mean, one overestimates the influence of the faster vehicles. If the influence of the faster vehicles on speeds is overestimated, the average speed u is overestimated (compared to the space mean speed u_m).

The discussion above might be conceived as academic. However, if we look at empirical data, then the differences between the time–mean speeds and space–mean speeds become apparent. Figure 7.4 shows an example where the time–mean speed and space–mean speed have been computed from motorway individual vehicle data collected at the A9 motorway near Amsterdam, the Netherlands. Figure 7.4 clearly shows that the differences between the speeds can be as high as 100%; for more details we refer to Knoop et al. (2009). Also note that the space–mean speeds are always lower than the time–mean speeds. Since in most countries where inductive loops are used to monitor traffic flow operations, *arithmetic–mean speeds* are computed and stored, average speeds are generally overestimated, affecting travel time estimations. Furthermore, since $q = ku$ can only be used for *space–mean speeds*, we cannot determine the density k from the local speed and flow measurements, complicating the use of the collected data from, e.g., traffic information and traffic management purposes.
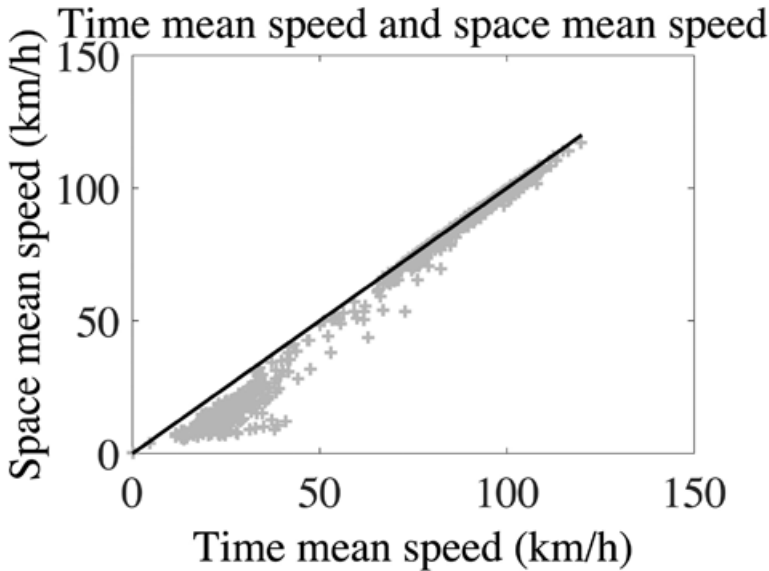
Source: Knoop (2020).

Figure 7.4   Differences between time–mean speed and space–mean speed for the A9 motorway

7.3.3   Generalized Traffic Flow Variables

Alternate measurement methods, such as automatic vehicle identification (AVI), radar, and floating-car measurements, provide new ways to determine the flow variables described above. One of the benefits of these new methods is that they provide information about the temporal and spatial aspects of traffic flow. For instance, using video we can observe the density in a region directly, rather than determine the density from local observations.

For the relation between instantaneous and local variables, the work of Edie (1965) is very relevant. Edie (1965) introduces generalized definitions of flow, density, and speed.

Consider a rectangular region in time and space with dimensions T and X respectively (see Figure 7.5). Let d_i denote the total distance travelled by vehicle i during period T and let r_i denote the total time spent in region X. Let us define the total distance travelled by all vehicles by:
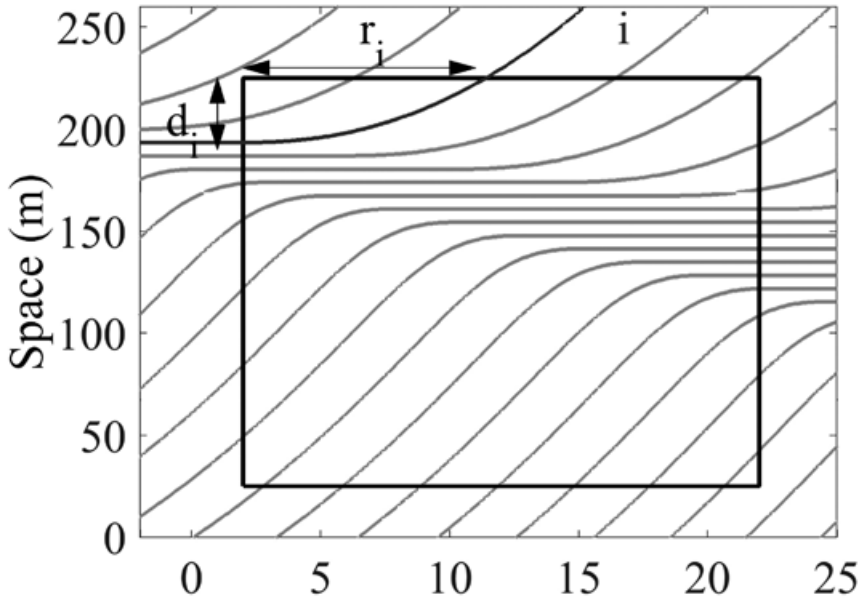
$$P = \sum_i d_i \tag{7.8}$$

Based on this quantity P, which is referred to as the *performance*, Edie defined the generalized flow as follows:

$$q = \frac{P}{XT} \tag{7.9}$$

Note that we can rewrite this equation as follows:

$$q = \frac{\sum_i d_i / X}{T} \tag{7.10}$$

Figure 7.5 Generalization of concepts of flow and density to a generic area in space-time

Let us now define the total travel time R as follows:

$$R = \sum_i r_i \tag{7.11}$$

Edie (1965) defines the generalized density by:

$$k = \frac{R}{XT} = \frac{\sum_i r_i / T}{X} \tag{7.12}$$

For the generalized speed, the following intuitive definition is used:

$$u = \frac{q}{k} = \frac{P}{R} = \frac{\text{total distance travelled}}{\text{total time spent}} \tag{7.13}$$

These definitions can be used for any region in space–time, even non-rectangular ones. These definitions apply to regions in time and space, and will turn out to be increasingly important with the advent of new measurement techniques. This is because traditionally traffic was

measured at one location (for instance, by loop detectors). New techniques (apps, connected vehicles) will measure at various locations and times, hence a measurement in space and time is necessary.

Moreover, new techniques (apps on phones) allow tracking for *some individual vehicles*, but do not show all vehicles. This gives completely different challenges to estimating traffic. Estimations of total number of vehicles that pass by still remains useful, as do derivative measures like flow, density, or speed. Van Erp et al. (2019) shows how these data, including findings on overtaking, can be used to get an overall view of the total traffic flow even though not all vehicles are being measured.

7.4   MICROSCOPIC AND MACROSCOPIC FLOW CHARACTERISTICS

The preceding sections introduced the different microscopic and macroscopic variables. This section shows the most common flow characteristics, entailing both relations between the flow variables, typical distribution, etc. These flow characteristics in a sense drive the traffic flow dynamics that will be discussed in this section. Next to providing a short description of the characteristics and their definition, the section will discuss empirical examples as well as key issues in identifying these parameters.

7.4.1   Headway Distributions

If we would collect headways at a specific location x, then we would observe that these headways are not constant but rather follow some probability distribution function. This is also the case when the flow is stationary during the data collection period. The causes are manifold: there are large differences in driving behaviour between different drivers, and differences in the vehicle characteristics, but there is also variation in the behaviour of one driver. A direct and important consequence of this is that the capacity of the road, which is largely determined by the driving behaviour, is not constant either, but a stochastic variable.

The headway distribution can be described by a probability density function (p.d.f.) $f(h)$. In the literature, many different kinds of distribution functions have been proposed, with varying success. It can be shown that if the flows are small – there are few vehicle interactions – the *exponential distribution* will be an adequate model. When the flows become larger, there are more interactions amongst the vehicles, and other distributions are more suitable. A good candidate in many situations is the *log-normal distribution*; we refer to Cowan (1975) for more details. In Hoogendoorn (2005), an overview is given of estimation techniques for the log-normal distributions in specific situations.

The main problem with these relatively simple models is that they are only able to represent available measurements but cannot be extrapolated to other situations. If, for instance, we are interested in a headway distribution for another flow level than the one observed, we need to collect new data and re-estimate the model.

To overcome this, so-called *composite headway models* have been proposed. The main characteristic of these models is that they distinguish between vehicles that are flowing freely and those that are constrained by the vehicle in front. Buckley (1968) was one of the first to propose these models, assuming that the headways of the free driving vehicles are exponentially distributed. He showed that the probability density function $f(h)$ of the observed headways h can be described by the following function:

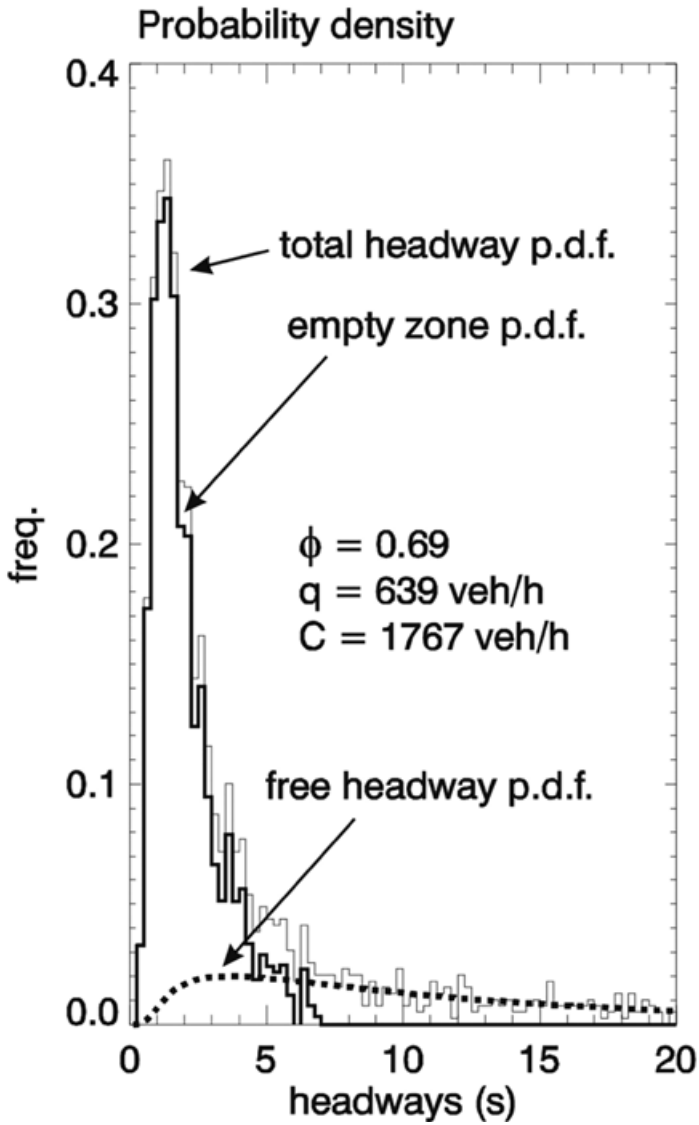$$f(h) \;=\; \phi g(h) + (1-\phi)w(h) \tag{7.14}$$

In this equation, g describes the probability density function of the headways of vehicles, which are following (also referred to as the distribution of the *empty zones*), while w denotes the probability density function of those vehicles that are driving freely. For the latter, an exponential distribution is assumed. ϕ denotes the fraction of vehicles that are following.

There are different ways to estimate these probability density functions from available headway observations. Wasielewski (1974), later improved by Hoogendoorn (2005), proposed an approach in which one does not need to choose a prior form of the constrained headway distribution. In illustration, Figure 7.6 shows an example of the application of this estimation method on a two-lane motorway in the Netherlands in the morning (the location is the so-called 'Doenkade').

This example nicely illustrates how the approach can be applied for estimating capacities, even if no capacity observations are available. We find the maximum flow (or capacity flow, C) when all drivers are following. We directly observe from the Buckley model (Buckley, 1968, see above) that the observed headways in that case follow g. The number of vehicles per unit of time is the inverse of time per vehicle. So the maximum flow equals 1 divided by the mean (minimum) headway value, under the condition that all vehicles are following. Accounting for the fact that we measure flows in vehicles per hour, and headways in seconds, we need to include a unit conversion. We then have that the capacity equals 3600 seconds per hour divided by the average headway (H, following distribution g) in seconds (or the expectation value thereof, indicated by E). We therefore get:

$$C \;=\; 3600/E(H) \text{ where } H \sim g \tag{7.15}$$

Using this approach, we can find estimates for the capacity even if there are no direct capacity observations available. For the example above, we can compute the mean empty zone value by looking at the p.d.f. $g(h)$, which turns out to be equal to 1.69. Based on this value, we find a capacity estimate of 3600/1.69 = 2134 vehicles per hour.

Probability density

Inside the figure:

- total headway p.d.f.
- empty zone p.d.f.
- $\phi = 0.69$
- $q = 639$ veh/h
- $C = 1767$ veh/h
- free headway p.d.f.

Axis labels: freq. (vertical), headways (s) (horizontal)

Source: Knoop (2020).

Note: In the figure, W denotes the gross headway, which is composed of the empty zone X and the free headway W-X

Figure 7.6   Composite headway probability density function (p.d.f.) fitted on data of the Doenkade site.

7.4.2 Desired Speed Distributions

Generally, the *free speed or desired speed of a driver–vehicle combination* (simply called vehicle or driver in the ensuing) is defined by the speed driven when other road users do not influence the driver. Knowledge of free speeds on a road under given conditions is relevant for a number of reasons. For instance, the concept of free speed is an important element in many traffic flow models. In illustration, the free speed distribution is an important input for many microscopic simulation models. Insights into free speeds and their distributions are also important from the viewpoint of road design, and for determining suitable traffic rules for a certain facility. For instance, elements of the network should be designed such that drivers using the facility can traverse the road safely and comfortably. It is also of interest to see how desired speed distributions change under varying road, weather and ambient conditions, and how these distributions vary for different types of travellers. That is, speed distribution is an important characteristic of the driver population for design issues.

The free speed will be influenced by the characteristics of the vehicle, the driver, the road, and (road) conditions such as weather and traffic rules (speed limits). Botma (1999) describes how individual drivers choose their free speed, discussing a behavioural model relating the free speed of a driver to a number of counteracting mental stresses a driver is subjected to. A similar model can be found in Jepsen (1998). However, these models have not been successful in their practical application. The problem of determining free speed distributions from available data is not trivial. In Botma (1999), an overview of alternative free speed estimation approaches is presented. Botma (1999) concluded that all methods he reviewed have severe disadvantages, which is the reason why another estimation approach is proposed. This approach is based on the concept of censored observations (Nelson, 1982) using a parametric estimation approach to estimate the parameters of the free speed distribution. Speed observations are marked as either censored (constrained) or uncensored (free flowing) using subjective criteria (headway and relative speed). Hoogendoorn (2005) presents a new approach to estimating the distribution of free speeds based on the method of censored observations.

7.4.3 Gap Acceptance and Critical Gaps

Gap acceptance is a process that occurs in different traffic situations, such as crossing a road, entering a roundabout, or performing an overtaking manoeuvre on a bi-directional road. The minimum gap that a driver will accept is generally called the critical gap. Mathematical representations of the gap acceptance process are an important part of, for instance, traffic simulation models.

In general terms the gap acceptance process can be described as follows: traffic participants that want to make a manoeuvre estimate the space they need and estimate the available space. Based on the comparison between required and available space, they decide to start the manoeuvre or postpone it. The term space is deliberately somewhat vague; it can be expressed either in time or in distance. The required space is dependent on characteristics of the traffic participant, the vehicle, and the road. The available space is dependent on the characteristics of, for instance, the on-coming vehicles and the vehicle to be overtaken (the passive vehicle).

Traffic participants have to perceive all these characteristics, process them, and come to a decision. Humans highly differ in perception capabilities, e.g., the ability to estimate distances can vary substantially between persons, and they differ in the acceptance of risk. The total acceptance process is dependent on many factors of which only a subset is observable. This has led to the introduction of stochastic models.

Many different methods to estimate the distribution of critical gaps, from observations of the gap acceptance process in reality, can be found in the literature (Brilon et al., 1999). Let us consider the problem of estimating the critical gap distribution. Suppose, as an example, a driver successively rejects gaps of 3, 9, 12 and 7 s and accepts a gap of 19 s. The only thing one can conclude from these observations is that this driver has a critical gap between 12 and 19 s. Stated in other words: the critical gap cannot be observed directly. The observations are, thus, censored. Note that it can also be concluded that only the *maximum of the rejected gaps* is informative for the critical gap (assuming that driver behaviour is consistent); the smaller gaps are rejected by definition.

7.4.4   Capacity and Capacity Estimation

Capacity is usually defined as follows: "The maximum hourly rate at which persons or vehicles can reasonably be expected to traverse a point or uniform section of a lane or roadway during a given time period (usually 15 minutes) under prevailing roadway, traffic, and control conditions" (National Academies of Sciences, Engineering, and Medicine, 2022).

Maximum flows (maximum free flows of queue discharge rates) are not constant values and vary under the influence of several factors. Factors influencing the capacity are, among other things, the composition of the vehicle fleet, the composition of traffic with respect to trip purpose, weather-, road-, and ambient conditions, etc. These factors affect the behaviour of driver vehicle combinations and thus the maximum number of vehicles that can pass a cross-section during a given time period. Some of these factors can be observed and their effect can be quantified. Some factors can, however, not be observed directly. Furthermore, differences exist between drivers implying that some drivers will need a larger minimum time headway than other drivers, even if drivers belong to the same class of users. As a result, the minimum headways will not be constant values but follow a distribution function (see discussion on headway distribution modelling). Observed maximum flows thus appear to follow a distribution. The shape of this distribution depends on, among other things, the capacity definition and measurement method or period. In most cases, a normal distribution can be used to describe the capacity.

Several researchers have pointed out the existence of two different maximum flow rates, namely pre-queue and queue discharge respectively (e.g., Cassidy and Bertini, 1999). Each of these has its own maximum flow distribution. We define the *pre-queue maximum flow* as the maximum flow rate observed at the downstream location just before the on-set of congestion (a queue) upstream. These maximum flows are characterized by the absence of queues or congestion upstream of the bottleneck, high speeds, instability leading to congestion on-set within a short period, maximum flows showing a large variance. The *queue discharge flow* is the maximum flow rate observed at the downstream location as long as congestion exists. These

maximum flow rates are characterized by the presence of a queue upstream of the bottleneck, lower speeds and densities, a constant outflow with a small variance which can sustain for a long period, however with lower flow rates than in the pre-queue flow state. Both capacities can only be measured downstream of the bottleneck location. The size of the drop depends on the bottleneck type, and particularly on the speed in the queue (Yuan et al., 2015). Average capacity drop changes are in the range of -1% to -15%.

There are many approaches that can be applied to compute the capacity of a specific piece of infrastructure. The suitability of the approach depends on a number of factors, such as:

1. the type of infrastructure (e.g., motorway without off- or on-ramps, on-ramp, roundabout, unsignalized intersection, etc.);
2. type of data (individual vehicle data, aggregate data) and time aggregation;
3. location of data collection (upstream of, in or downstream of the bottleneck);
4. traffic conditions for which data are available (congestion, no congestion).

We refer to Minderhoud et al. (1996) for a critical review of approaches that are available to estimate road capacity.

7.4.5   Fundamental Diagrams

The fundamental diagram describes a statistical relation between the macroscopic traffic flow variables: flow, density and speed. There are different ways to represent this relation, but the most often used is the relation $q = Q(k)$ between the flow and the density. Using the continuity equation, the other relations $u = U(k)$ and $u = U(q)$ can be easily derived.

To understand the origin of the fundamental diagram, we can interpret the relation from a driving behaviour perspective. To this end, recall that the flow and the density relate to the (average) time headway and distance headway according to Equations 7.4 and 7.5 respectively. Based on this, we can clearly see which premise underlies the existence of the fundamental diagram: *under similar traffic conditions, drivers will behave in a similar way.* That is, when traffic operates at a certain speed u, then it is plausible that (on average) drivers will maintain (on average) the same distance headway $s = 1/k$. This behaviour – and therewith the relation between speed and density – is obviously dependent on factors like weather, road characteristics, the composition of traffic, traffic regulations, and so forth.

Figure 7.7 shows typical examples of the relation between flow, density, and speed. The figure shows the most important points in the fundamental diagram, which are the *roadway capacity C*, the *critical density k_c* and the *critical speed u_c* (the density and speed occurring at capacity operations), the *jam density k_{jam}* (density occurring at zero speed), and the *free speed u_0*. In the figure, we clearly see the difference between the free conditions ($k < k_c$) and the congested conditions ($k > k_c$).

It is tempting to infer causality from the fundamental diagram. It is often stated that the relation $u = U(k)$ describes the fact that with increasing density (e.g., reduced spacing between vehicles), the speed is reducing. It is, however, more the other way around. If we take a driving behaviour perspective, then it seems more reasonable to assume that with reduced speed of the leader, drivers need smaller distance headways to drive safely and comfortably.
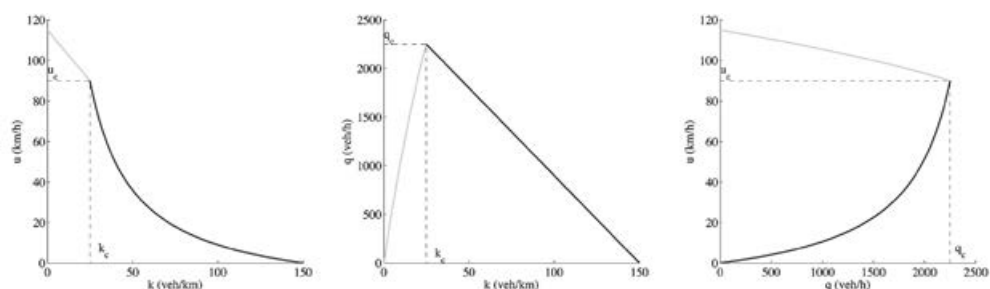
Figure 7.7   Example of fundamental diagram

Fundamental diagrams are often determined from real-life traffic data. This is usually done by assuming that stationary periods can be identified during data measurements. To obtain meaningful fundamental diagrams, the data collection must be performed at the correct location during a selected time period.

7.5   DEVELOPMENTS FOR THE FUTURE: CONNECTED AND AUTOMATED VEHICLES

In the past decade, much effort has been put into the automation of vehicles. The driving tasks of drivers can be handed over to the vehicle. In this section, we will describe the levels of automation, how this will interfere with the traffic operations, and the potential effects.

For traffic automation, various levels of autonomy have been defined. The most commonly used reference is the SAE levels of automation (SAE, 2021). Here, we report their defined levels in a simplified way. For the full description, we refer to (SAE, 2021). There are six levels of autonomy defined, i.e., level 0 to 5.

Level 0 is the automation level where the driver is performing the driving tasks. He is (almost) continuously giving input to the vehicle in terms of steering (lateral driving task) and determining the vehicle's speed (longitudinal driving task).

In level 1, the driver is supported in one driving task, lateral or longitudinal. He either needs to continuously steer or determine the speed. The other dimension is taken over by the system, by means of a system that keeps the vehicle in its lane or determines the acceleration automatically by means of adaptive cruise control (ACC).

In level 2 the car can accelerate and decelerate by means of an ACC and keep itself in lanes. The driver can drive "hand and feet off". Yet, the driver needs to constantly monitor the environment and the system, and the driver should be able to take over control instantaneously without any warning. As of 2022, there are vehicles available that can achieve level 2 automation.

In level 3, the vehicle also monitors the system and warns the driver if he needs to take over ("fallback"). The driver should be able to take over when the system demands so. As of April 2022, no vehicles are on the market that have this level of automation. Some manufacturers claim their vehicles can reach level 3, but at present it has not been legally approved.

From level 4 onwards, this fallback is taken over by the system and the car will – within its predefined bounds, or "operational design domain" – not need driver input. The system is designed to work within a certain operational design domain. What this operational design domain is can be fully determined by the car manufacturer. Examples could be a speed range (e.g., between 20 and 60 km/h), a specific road (e.g., the M1 motorway, between kilometre 12 and 15 in the left lane), or weather conditions (daylight, no precipitation, good sight), or combinations of these. As long as the vehicle remains within this operational domain, in level 4 automation, the driver has no need to interfere with the vehicle and does not need to be the fallback option. This changes if the vehicle comes out of the operational design domain.

For level 5, these operational design domains are removed. Level 5 automated vehicles can drive themselves without human supervision anywhere and anytime.

Note here that the step from level 2 to level 3 implies a different driving experience where a driver (within the operational design domain) does not need to pay constant attention to the road.

An interesting point is that from level 2 onwards, for vehicles that are on the market now, the "operational" driving is done by the vehicle itself. Indeed, the driver needs to constantly monitor and intervene directly and whenever necessary. However, with good systems, this will be rare, and the vehicle decides on its speed and lane itself. The analyses of driving behaviour and how this influences the traffic stream should from this level onwards, therefore, (also) be performed for the vehicle.

In early studies of automated vehicles, it was expected that vehicle automation would strongly increase road capacity. Namely, automated systems have no reaction time, and vehicles can travel closely together at high speed. This will lead to short time headways and hence a high capacity. Rao et al. (1993) predicted flows of up to 6900 veh/h/lane. On a vehicle level, that would mean a time headway of approximately 0.5 seconds. Once the (level 1 and level 2 automation) ACC systems became more commonly available, real-world tests have been performed with vehicles that are on the market. Early ACC systems were studied by Milanés and Shladover (2014). They concluded that the capacity actually decreased because the systems keep a longer time headway than human drivers. This could be in line with reports on lower roadway capacities, which is a change compared to a trend of decades of increasing road capacity for the same road layout (Shiomi et al., 2019; Knoop and Hoogendoorn, 2022).

The large headways found for one brand in the early days have been confirmed for more recent vehicles and a variety of brands (Knoop et al., 2019). The main conclusion of this study is that ACC systems make the traffic stream unstable. That is, if the first vehicle brakes, the (ACC equipped) follower brakes stronger, and the next (ACC equipped) follower even stronger, etc. In such a way, small perturbations will grow to traffic jams and potentially dangerous situations. The market is currently (2022) being flooded with more vehicles with ACC systems, and potentially with updates of current systems. The response of ACC systems to a disturbance is empirically studied, and a database is built where these data are stored and made accessible for researchers; see Ciuffo et al. (2020).

Note that the previous section discussed autonomous vehicles, i.e., vehicles that drive autonomously (i.e., by themselves, without influence from others). This is different from vehicles that exchange messages with each other or a control centre; the latter are called

connected vehicles. A common combination for studies of the impact of new technologies is autonomous connected vehicles, which some researchers refer to as automated vehicles. Some of the drawbacks of autonomous vehicles, like the instabilities, can be overcome by sending messages. In the future, communication on anticipated braking manoeuvres between vehicles can potentially solve this stability issue. If a vehicle is certain that it will get a timely message before it needs to brake, it can drive closer to its predecessor and still there would be no need to over-react, or it can form platoons with other vehicles to cross traffic lights.

7.6   TRAFFIC FLOW DYNAMICS AND SELF-ORGANIZATION

So far, we have discussed the main microscopic and macroscopic characteristics of traffic flow. In doing so, we have focused on the static characteristics of traffic flow. However, there are different characteristics, which are dynamic in nature, or rather, have to do with the dynamic properties of traffic flow.
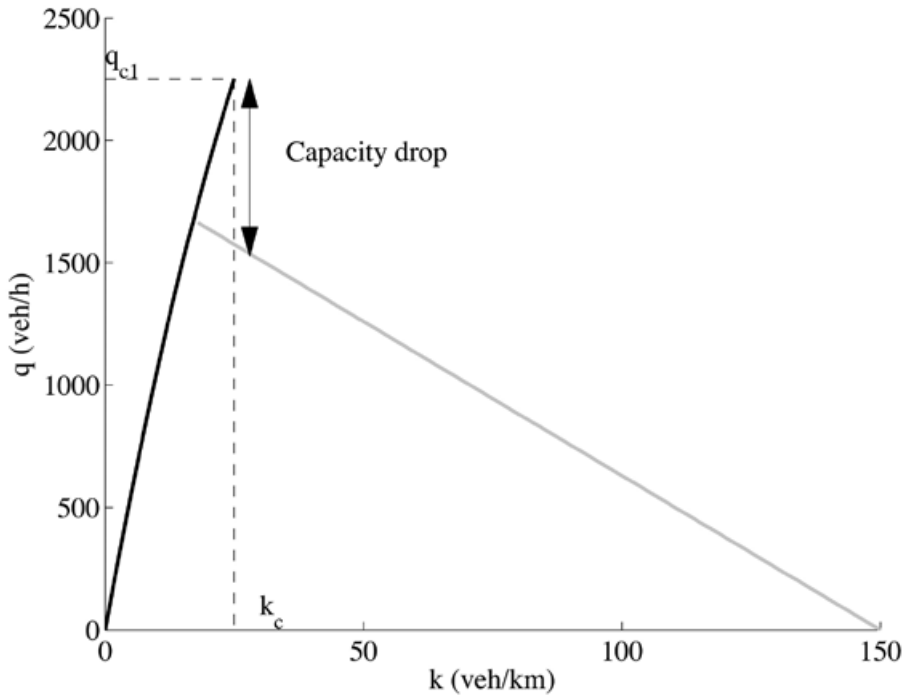
7.6.1   Capacity Drop

The first phenomenon that we discuss is the so-called capacity drop. The capacity drop describes the fact that once congestion has formed, drivers are not maintaining a headway as close as before the speed breakdown. Therefore, the road capacity is lower. This effect is considerable, and values of a reduction up to 30% are quoted (Hall and Agyemang-Duah, 1991; Cassidy and Bertini, 1999; Chung et al., 2007; Yuan et al., 2015). The effect of the capacity drop is illustrated in Figure 7.8. Causes of the capacity drop lie in the individual driving behaviour. The exact cause is unknown and might lie in lane changing (Laval and Daganzo, 2006) or car-following/acceleration behaviour (Yuan et al., 2017).

7.6.2   Traffic Hysteresis

The different microscopic processes that constitute the characteristics of a traffic flow take time: a driver needs time to accelerate when the vehicle in front drives away when the traffic signal turns green. When traffic conditions on a certain location change, for instance when the head of a queue moves upstream, it will generally take time for the flow to adapt to these changing conditions.

Generally, however, we may assume that given that the conditions remain unchanged for a sufficient period of time – say, five minutes – traffic conditions will converge to an average state. This state is often referred to as the equilibrium traffic state. When considering a traffic flow, this equilibrium state is generally expressed in terms of the fundamental diagram. That is, when considering traffic flow under stationary conditions, the flow operations can – on average – be described by some relation between speed, density, and flow. This is why the speed–density relation is often referred to as the equilibrium speed.

From real-life observations of traffic flow, it can be observed that many of the data points collected are not on the fundamental diagram. While some of these points can be explained by

Figure 7.8   The capacity drop in the flow density diagram

stochastic fluctuations (e.g., vehicles have different sizes, drivers have different desired speeds and following distances), a part of it can be structural, and stem from the dynamic properties of traffic flow. That is, they reflect so-called transient states (i.e., changes from congestion to free flow (acceleration phase) or from free flow to congestion (deceleration phase) of traffic flow. It turns out that generally, these changes in the traffic state are not on the fundamental diagram. In other words: if we consider the average behaviour of drivers (assuming stationary traffic conditions), observed mean speeds will generally not equal the "equilibrium" speed. The term "equilibrium" reflects the fact that the observed speeds in time will converge to the equilibrium speed, assuming that the average conditions remain the same. That is, the average speed does not adapt instantaneously to the average or equilibrium speed.

   This introduces traffic hysteresis, i.e., at the same time, drivers keep a different headway speed during acceleration then during deceleration. Figure 7.9 shows the first empirical observation thereof by Treiterer and Myers (1974). The figure shows the time it takes for a platoon to pass a point along the roadway. The longer the arrow is, the longer that time is, and hence the lower the flow (vehicles/hour). The arrow is long at the beginning since some drivers are not car-following yet. At the second arrow, all vehicles are car-following and the flow is high (short arrow). In the disturbance, the flow is very low, and we find a long arrow. After the disturbance, the flow increases, but the headways are longer than before the vehicles entered the disturbance. Note that also in exiting the traffic jam, all vehicles will be in car-following mode.
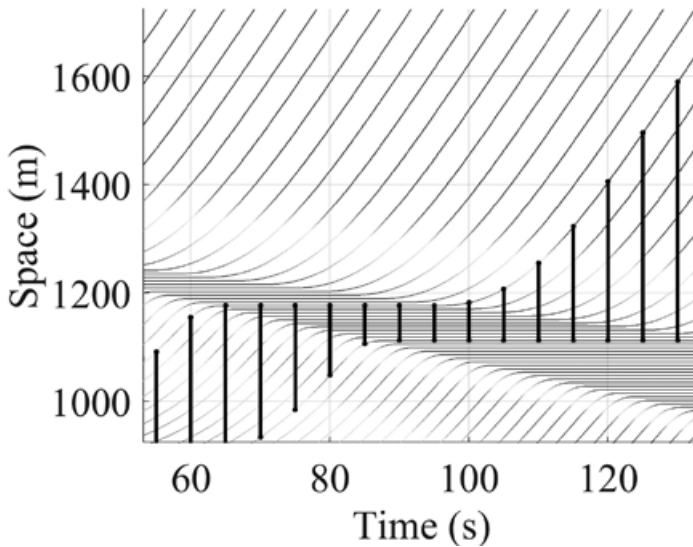
Source: Treiterer and Myers (1974).

Figure 7.9   Vehicle trajectories collected from airborne platform clearly showing differences in average platoon length before and after disturbance

Decades later, it has been realized that the hysteresis is perhaps not best measured at one time instance. Whereas hysteresis still occurs, effects might change or have a different magnitude as shown by Laval (2011).

7.6.3   Three Phase Traffic Flows, Phase Transitions, and Self-Organization

Traffic can have different states or phases. In the early 2000s there was a strong debate on the number of phases and phase transitions. Kerner (2004) commented that there are three phases (free flow, synchronized flows, and jams), whereas many others (e.g., Treiber et al., 2000) argued there are two (free flow and congestion). In traffic patterns, we can identify congestion with different speeds and different causes. There can be freely flowing traffic at high speed, a traffic moving at lower speed caused by a restriction of capacity, or completely stopped traffic. These are identifiable in traffic and are consistent with the states Kerner has distinguished.

   If the vehicles indeed come to a complete stop, arriving vehicles will need to stop upstream. It will hence grow at the upstream end. At the same time, at the downstream end, vehicles might start moving again, causing a backward moving front of the queue. As a pattern, this queue therefore travels in time over space, as is visible in Figure 7.9. This pattern is called a stop-and-go wave, or (by Kerner) a wide moving jam. The wave speed is approximately 18 km/h opposite the driving direction.

The causes of these jams differ. An example is a jam caused by a bottleneck, such as an on-ramp. In this situation, the simple fact that traffic demand is at some point in time larger than the rest capacity (being the motorway capacity minus the inflow from the on-ramp) causes a jam. Note that these kinds of phase transitions can be described by basic flow theories and models (shockwave theory, kinematic wave models) adequately. As an additional remark, note that these transitions are, although induced, still random events since both the free flow capacity and the supply are random variables.

However, not all phase transitions are induced (directly); some are caused by intrinsic ("spontaneous") properties of traffic flow. An example is the transition from a jam with moving traffic to a stop-and-go wave. (referred to by Kerner as wide moving jams). Due to the unstable nature of specific denser traffic (in the congested state of the fundamental diagram), small disturbances in the congested flow will grow from one vehicle to its leader, and hence also over time. If the gaps between platoons of vehicles are not large enough to absorb a disturbance, it can cause traffic to come to a complete stop (traffic instability; Pueboobpaphan and Van Arem, 2010).

This phenomenon is quite common in day-to-day motorway traffic operations. Figure 7.10 shows an example of the A4 motorway in the Netherlands. Figure 7.10 shows an example of the A4 motorway in the Netherlands. A bottleneck can be identified around km 55. One can find stop-and-go waves propagating backwards at approximately 18 km/h. Note that as wide
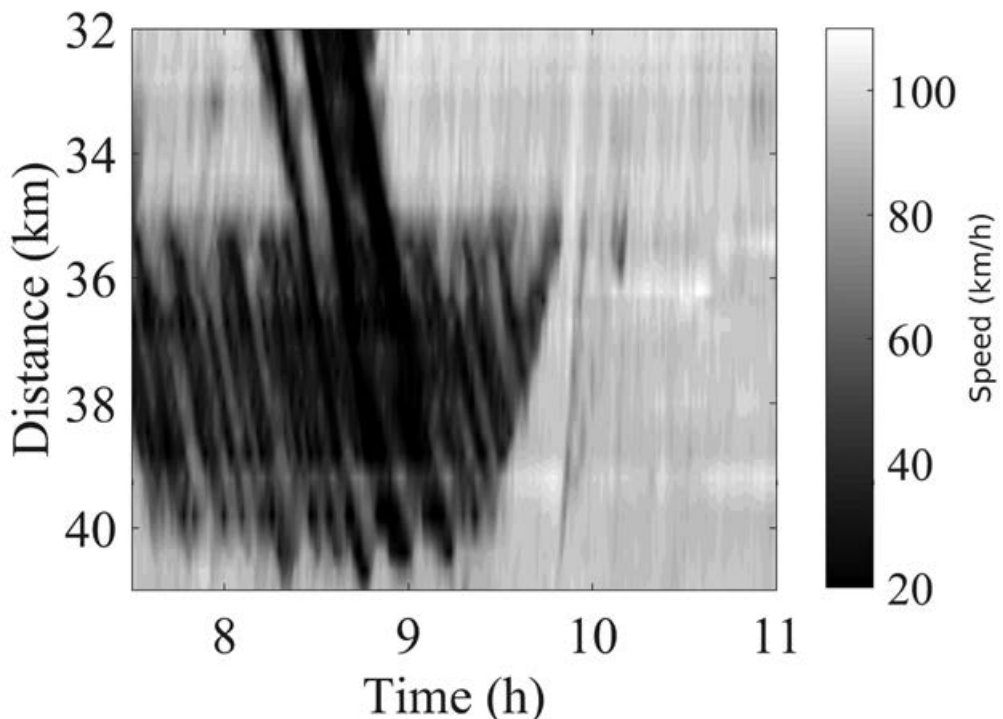

Figure 7.10   Typical traffic patterns on the A4 motorway in the Netherlands

moving jams. Note that as wide moving jams have an outflow rate which is about 30% lower than the free flow capacity, these jams are actually quite undesirable from a traffic efficiency perspective. Furthermore, they imply additional braking and acceleration, yielding increased fuel consumption and emission levels.

7.7   MULTI-LANE TRAFFIC FLOW FACILITIES

Up to now, the chapter has considered each lane of the freeway to be equal. However, there are considerable differences between them. This section only introduces a basic concept. For more insight we refer to the literature mentioned in this paragraph. For the sake of simplicity, we assume here right-hand driving. For countries where a left-hand driving rule applies, like Japan, the United Kingdom, or Australia, the lanes are exactly opposite. Daganzo (2002a, 2002b) poses a theory classifying the drivers as *slugs*, defined by their low desired speed, and *rabbits*, defined by their higher desired flow speed. He states that as soon as the speed in the right lane will go under a threshold, rabbits will move to faster lanes at the left. The theory further states that even if the density in the right lane is lower than in the left lane, the rabbits will not change towards the right lane as long as the speeds in the left lane are higher. This traffic state, with two different speeds, is called a *two pipe regime*, since traffic is flowing as it were in two different, unrelated pipes. In this state, there is no equal density in both lanes. Only once the density in the left lane increases that much that the speed decreases to a value lower than the speed in the right lane, will rabbits move towards the right lane. Then, the rabbits will redistribute themselves in such a way that the traffic in both lanes flows at the same speed. This is called a *one pipe regime*.

Note that the speeds in different lanes at the same densities can be different, due to these effects or basically due to the driver population on that lane. This leads to different fundamental diagrams in the left and right lanes. Usually, the free flow speed in the left lane is higher than in the right lane, due to the higher fraction of rabbits in that lane (see Knoop et al., 2010). Kerner (2004) poses a similar theory on multi-lane traffic flow facilities. This unbalanced lane flows cause inefficiencies in road traffic. By actively influencing lane choice or lane changes, one can try to increase road capacity and hence reduce delays. This has been an active field of research in the past decade (e.g., Roncoli et al., 2017, and Nagalur Subraveti et al., 2020).

7.8   MICROSCOPIC FLOW MODELS

Traffic flow models can be used to simulate the traffic, for instance, to evaluate *ex-ante* the use of a new part of the infrastructure. Traffic flow models may be categorized using various dimensions (deterministic or stochastic, continuous or discrete, analytical or simulation, and so forth). The most common classification is the distinction between microscopic and macroscopic traffic flow modelling approaches. However, this distinction is not unambiguous,

due to the existence of hybrid models. This is why here models are categorized based on the following aspects:

1. **Representation** of the traffic flow in terms of flows (macroscopic), groups of drivers (macroscopic), or individual drivers (microscopic).
2. **Underlying behavioural theory**, which can be based on characteristics of the flow (macroscopic), or individual drivers (microscopic behaviour).

The remainder of this section uses this classification to discuss some important flow models. Table 7.1 depicts an overview of these models.

Table 7.1   Overview of traffic flow model classification

Representation	Behavioural rules	
	Microscopic	Macroscopic
Vehicle based	Microscopic flow models	Particle models
Flow based	Gas-kinetic models	Macroscopic models

The observed behaviour of drivers, i.e., headways, driving speeds, driving lane, is influenced by different factors, which can be related to the driver–vehicle combination (vehicle characteristics, driver experience, age, gender, and so forth), the traffic conditions (average speeds, densities), infrastructure conditions (road conditions), and external situational influences (weather, driving regulations). Over the years, different theories have been proposed to (dynamically) relate the observed driving behaviour to the parameters describing these conditions.

In doing so, different driver subtasks are often distinguished. In general, two types of driver tasks are distinguished: longitudinal tasks (acceleration, maintaining speed, distance keeping relative to leading vehicle) and lateral tasks (lane changing, overtaking). In particular the longitudinal and (to a lesser extent) the lateral interaction subtasks have received quite some attention in traffic flow theory research.

A microscopic model provides a description of the movements of individual vehicles that are considered to be a result of the characteristics of drivers and vehicles, the interactions between driver–vehicle elements, the interactions between driver–vehicle elements and the road characteristics, external conditions and the traffic regulations and control. Most microscopic simulation models assume that a driver will only respond to the one vehicle that is driving in the same lane, directly in front of him (the leader).

When the number of driver–vehicle units on the road is very small, the driver can freely choose his speed given his preferences and abilities, the roadway conditions, curvature, prevailing speed limits, and so forth. In any case, there will be little reason for the driver to adapt his speed to the other road users. The target speed of the driver is the so-called free speed. In real life, the free speed will vary from one driver to another, but also the free speed of a single driver will change over time. Most microscopic models assume however that the free speeds have a constant value that is driver-specific. When traffic conditions deteriorate, the driver will no longer be able to choose the speed freely, since he will not always be able to overtake or pass a slower vehicle. The driver will need to adapt his speed to the prevailing traffic conditions, i.e., the driver is following. In the remainder of this section, we will discuss some of these

car-following models. Models for the lateral tasks, such as deciding to perform a lane change and gap acceptance, will not be discussed in this section in detail. The Minimizing Overall Braking Induced by Lane Changes (MOBIL) model (Kesting et al., 2007) or Lane Change Model with Relaxation and Synchronization (LMRS) (Schakel et al., 2012) models provide a good basis for realistic lane changing.

7.8.1   Safe-Distance Models

The first car-following models were developed by Pipes (1953) and were based on the assumption that drivers maintain a safe distance. A good rule for following vehicle *i*-1 at a safe distance s_i is to allow at least the length S_0 of a car between vehicle *I* and a part which is linear with the speed v_i at which *i* is travelling:

$$s_i = S(v_i) = S_0 + T_r v_i \tag{7.16}$$

Here, S_0 is the effective length of a stopped vehicle (including additional distance in front), and T_r denotes a parameter (comparable to the reaction time). A similar approach was proposed by Forbes et al. (1958). Both Pipes' and Forbes' theory were compared to field measurements. It was concluded that according to Pipes' theory, the minimum headways are slightly less at low and high velocities than observed in empirical data. However, considering the models' simplicity, agreement with real-life observations was amazing (Pignataro, 1973).

7.8.2   Stimulus-Response Models

However, safe-distance models do not seem to capture many phenomena observed in real-life traffic flows, such as hysteresis, traffic instabilities, etc. Stimulus response models are dynamic models that describe the reaction of drivers as a function of changes in distance, speeds, etc., relative to the vehicle in front, more realistically, e.g. by considering a finite reaction time. These models are applicable to relatively busy traffic flows, where the overtaking possibilities are small, and drivers are obliged to follow the vehicle in front of them. Drivers do not want the gap in front of them to become too large, so that other drivers can enter it. At the same time, the drivers will generally be inclined to keep a safe distance.

Stimulus response models assume that drivers control their acceleration (*a*). The well-known model of Chandler et al. (1958) is based on the intuitive hypothesis that a driver's acceleration is proportional to the relative speed $v_{i-1} - v_i$:

$$a_i(t) = \frac{d}{dt} v_i(t) = \alpha \left(v_{i-1}(t - T_r) - v_i(t - T_r) \right) \tag{7.17}$$

where T_r again denotes the overall reaction time, and α denotes the sensitivity. Based on field experiments, conducted to quantify the parameter values for the reaction time T_r and the sensitivity α, it was concluded that α depended on the distance between the vehicles: when the vehicles were close together, the sensitivity was high, and vice versa.

Stimulus-response models have been mainly applied to single lane traffic (e.g., tunnels, cf. Newell, 1961) and traffic stability analysis (Herman, 1959; May, 1990). It should be noted that

no generally applicable set of parameter estimates has been found so far, i.e., estimates are site-specific. An overview of parameter estimates can be found in Brackstone and McDonald (1999).

7.8.3   Psycho-Spacing Models

The two car-following models discussed so far have a mechanistic character. The only human element is the presence of a finite reaction time T_r. However, in reality a driver is not able to:

1.   observe a stimulus lower than a given value (perception threshold);
2.   evaluate a situation and determine the required response precisely, for instance due to observation errors resulting from radial motion observation;
3.   manipulate the gas and brake pedal precisely.

Furthermore, due to the need to distribute his attention to different tasks, a driver will generally not be permanently occupied with the car-following task. This type of consideration has inspired a different class of car-following models, namely the *psycho-spacing models*. Michaels (1963) provided the bases for the first psycho-spacing based on theories borrowed from perceptual psychology; cf. Leutzbach and Wiedemann (1986).

The so-called action point models (an important psycho-spacing model) form the basis for a large number of contemporary microscopic traffic flow models. An attempt to put these effects into models has been made by Hoogendoorn et al. (2010).

7.9   MACROSCOPIC TRAFFIC FLOW MODELS

In the previous section we have discussed different microscopic traffic flow modelling approaches. In this section, we will discuss the main approaches that have been proposed in literature taking a macroscopic perspective.

7.9.1   Deterministic and Stochastic Queuing Theory

The most straightforward approach to model traffic dynamics is probably the use of queuing theory. In queuing theory, we keep track of the number of vehicles in a queue (n). A queue starts whenever the flow to a bottleneck is larger than the bottleneck capacity, where the cars form a virtual queue. The outflow of the queue is given by the infrastructure (it is the outflow capacity of the bottleneck, given by C), whereas the inflow is the flow towards the bottleneck (q) as given by the traffic model. In an equation, this is written as:

$$dn \; = \; q(t)\,dt - C(t)\,dt \qquad\qquad (7.18)$$

The number of vehicles in the queue (n; *dn* stands for the change in number of vehicles in the queue) will evolve in this way until the queue is completely disappeared. Note that both the inflow and the capacity are time dependent in the description. For the inflow, this is due to the random distribution pattern of the arrivals of the vehicles. Vehicles can arrive in platoons

or there can be large gaps in between two vehicles. Also, the capacity is fluctuating. On the one hand, there are vehicle-to-vehicle fluctuations. For instance, some drivers have a shorter reaction time, hence a shorter headway leading to a higher capacity. On the other hand, the capacities will also on a larger scale depend on road or weather conditions (e.g., wet roads, night-time).

Figure 7.11 shows how the number of vehicles in the queue, n, fluctuates with time for a given inflow and outflow curve.
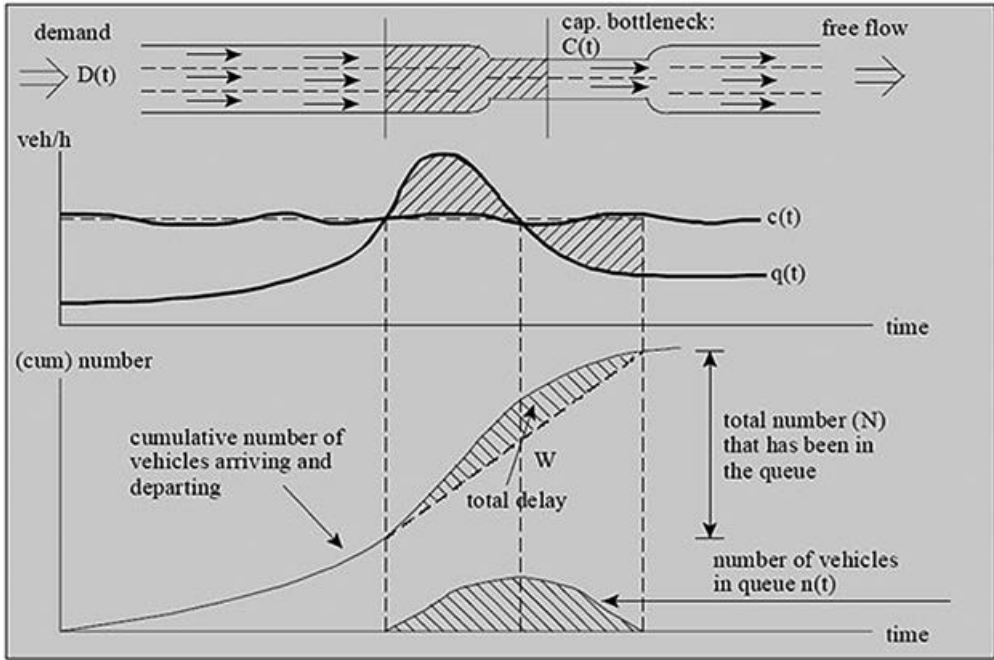

Figure 7.11 Functioning of queuing theory

The disadvantage of the queuing theory is that the queues have no spatial dimension, and they do not have a proper length either (they do not occupy space). Other models, which overcome these problems, are discussed below.

7.9.2 Shockwave Theory

Queuing theory provides some of the simplest models that can be used to model traffic flow conditions. However, in particular the spatial dimension of traffic congestion is not well described, or – in case of vertical queuing models – not described at all. *Shockwave theory* is able to describe the spatio-temporal properties of queues more accurately. This section briefly introduces shockwave theory.

A *shockwave* describes the boundary between two traffic states that are characterized by different densities, speeds, and /or flow rates. Shockwave theory describes the dynamics of

shockwaves, in other words: how does the boundary between two traffic states move in time and space.

Suppose that we have two traffic states: state 1 and 2. Let S denote the wave that separates these states. The speed of this shockwave S can be computed by:

$$w_{12} = \frac{q_2 - q_1}{k_2 - k_1} \qquad (7.19)$$

In other words, the speed of the shockwave equals the jump in the flow over the wave divided by the jump in the density. This yields a nice graphical interpretation (Figure 7.12): if we consider the line that connects the two traffic states 1 and 2 in the fundamental diagram, then the slope of this line is exactly the same as the speed of the shock in the time–space plane.
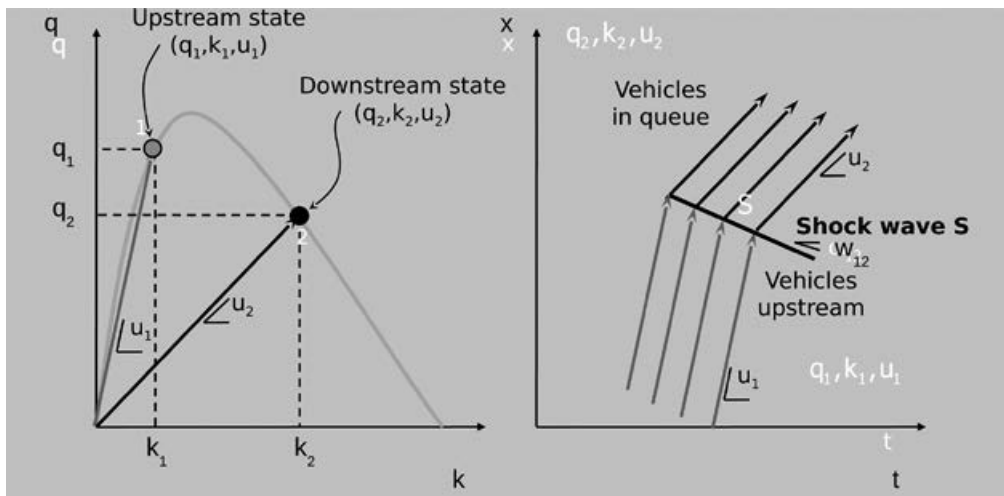
Figure 7.12   Graphical interpretation of shockwave speed

Shockwave theory provides simple means to predict traffic conditions in time and space. These predictions are largely in line with what can be observed in practice, but it has its limitations:

1. Traffic driving away from congestion does not smoothly accelerate towards the free speed but keeps driving at the critical speed.
2. Transition from the one state to the other always occurs jump-wise, not taking into account the bounded acceleration characteristics of real traffic.
3. No consideration of hysteresis.
4. No spontaneous transitions from the one state to the other.
5. Location of congestion occurrence is not in line with reality.

As a result, more advanced approaches have been proposed. To deal with this, continuum traffic models have been developed.

7.9.3   Continuum Traffic Flow Models

Continuum traffic flow deal with traffic flow in terms of aggregate variables, such as flow, densities, and mean speeds. Usually, the models are derived from the analogy between vehicular flow and flow of continuous media (e.g., fluids or gases), complemented by specific relations describing the average macroscopic properties of traffic flow (e.g., the relation between density and speed). Continuum flow models generally have a limited number of equations that are relatively easy to handle.

Most continuum models describe the dynamics of density $k = k(x,t)$, mean instantaneous speed $u = u(x,t)$, and the flow $q = q(x,t)$. The density $k(x,t)$ describes the *expected number of vehicles per unit length* at instant t. The flow $q(x,t)$ equals the *expected number of vehicles* flowing past cross-section x during per time unit. The speed $u(x,t)$ equals the *mean speed of vehicle* defined according to $q=ku$. For an overview of continuum flow models, we refer to Van Wageningen-Kessels et al. (2015).

7.10   NETWORK DYNAMICS

In the preceding sections, we have presented some of the main traffic flow characteristics. Using the microscopic and macroscopic models discussed, flow operations on simple infrastructure elements can be explained and predicted. To predict flow operations in a network is obviously more involved since it requires also predicting the route traffic demand profiles, which in turn means modelling route choice, departure time choice, mode choice, etc.

Interestingly, it turns out the *overall dynamics* of a traffic network can be described using a remarkably simple relation, referred to as the *macroscopic* or *network fundamental diagram (NFD)*. This diagram relates the vehicle accumulation – or average vehicle density – to the network performance. The network performance is defined by the flow, weighted by the number of lanes, and the length of the roadway segment for which the measured flow is representative.

This relation, which will be discussed in the following sections, shows one of the most important properties of network traffic operations, namely that its performance decreases when the number of vehicles becomes larger. In other words, when it is very busy in the network, performance goes down and less vehicles are able to complete their trip per unit of time. As a consequence, problems become even bigger.

7.10.1   Macroscopic Fundamental Diagram

Vehicular traffic network dynamics are atypical. Contrary to many other networks, network production (average rate at which travellers complete their trip) deteriorates once the number of vehicles in the network has surpassed the critical accumulation. Pioneering work of Daganzo and Geroliminis (2008) shows the existence of the NFD, clearly revealing this fundamental property. Figure 7.13 shows an example of the NFD. Knowledge of this fundamental property and its underlying mechanisms is pivotal in the design of effective traffic management.
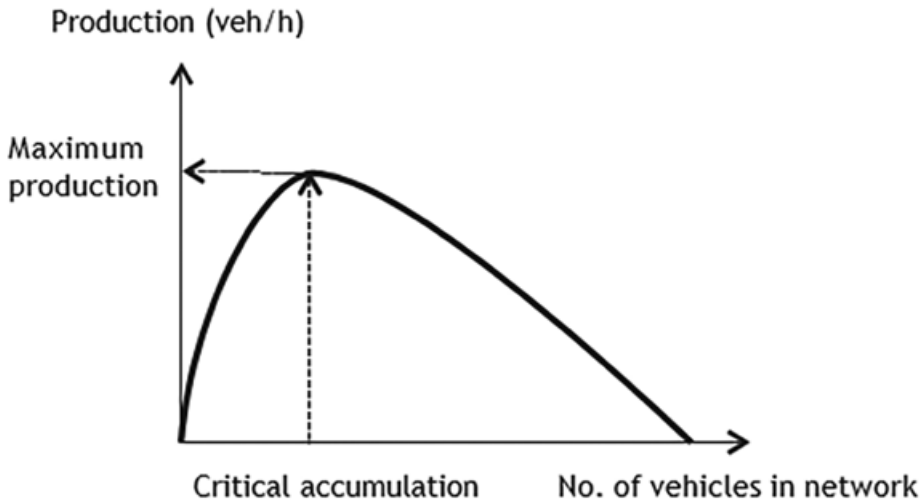
Figure 7.13   Example Network Fundamental Diagram

Developing a macroscopic description of traffic flow is not a new idea. Thomson (1967) found the relationship between average speed and flow using data collected streets in central London. Wardrop (1968) stated that this relation between average speed and flow decreased monotonically and Zahavi (1972) enriched Wardrop's theory by analysing real traffic data collected from various cities in the United Kingdom and the United States. Geroliminis and Daganzo (2008) have proven that NFDs exist in small networks, revealing the relationship between the outflow and accumulation in the network. This has later been confirmed for many other cities (Loder et al., 2019). The accumulation is the number of vehicles in the network. The outflow is also called trip completion rate, reflecting the rate at which trips reach their destinations. Similar to a conventional link fundamental diagram relating the local flow and density, three states are demonstrated on a NFD. When only a few vehicles use the network, the network is in the free flow condition and the outflow is low. With the increase of the number of vehicles, the outflow rises up to the maximum. Like the critical density in a link fundamental diagram, the value of corresponding accumulation when maximum outflow is reached is also an important parameter, called "sweet spot".

   As the number of vehicles further increases, travellers will experience delay. If vehicles continue to enter the network, it will result in a congested state where vehicles block each other and the outflow declines (congested conditions). Daganzo et al. (2011) illustrated that once a zone becomes only slightly too full, its performance can quickly deteriorate. In terms of control concepts, Keyvan-Ekbatani et al. (2012) developed rules for limiting the number of vehicles inside a zone in order not to exceed a critical number.

7.10.2   Causes for Network Degeneration

The two main causes for the production deterioration of overloaded networks are spill-back of queues possibly resulting in gridlock effects, and the capacity drop. Spill-back occurs due to the

simple fact that queues occupy space: a queue occurring at a bottleneck may propagate so far upstream that it will affect traffic flows that do not have to pass the bottleneck, e.g., when the queue passes a fork or an intersection upstream of the active bottleneck. As a result, congestion will propagate over other links of the network, potentially causing gridlock phenomena. The capacity drop describes the fact that the free flow freeway capacity is considerably larger than the queue discharge rate.

Vehicle automation can partially reduce the effects of network degradation. Traffic light settings can be further optimized to further stretch the network capacity (e.g., Keyvan-Ekbatani et al., 2019). Road capacity can be influenced by the vehicle type, and so can the capacity drop, for instance by pre-notifications (see Section 7.5). Routing might be another way to reduce the network degradation due to spill-back. Once, or even before, queues start to form, vehicles can be rerouted in order to avoid the growth of traffic jams. This means traffic that does not need to pass the bottleneck is not delayed by spill-back queues and can continue. This does come at a cost of inefficiency due to detours. In a fully connected system, these costs can be balanced.

7.11 BICYCLE TRAFFIC

Research into bicycle traffic has increased in the past decade. Whereas there are many studies on the choice of mode, the actual description of the way people cycle operationally has not taken off until recently.

For as long as the modelling assumptions hold, all tools and analyses as described above for vehicles can be used also for cycling. A main difference compared to vehicular traffic is the fact that cyclist traffic is not lane bound, and hence cyclists have freedom to choose a lateral position. Let us in brief revisit the elements in this chapter and indicate how these elements change in cycling traffic. If cyclists move along streams, all tools can be the same as for car traffic. Sometimes, no lanes are indicated, and they mix completely with other traffic, for instance in so-called *shared space.*

Variables can be defined in the same way for cyclists. Microscopically, one has only headways if one can define a leader. Edie's (1965) definitions (equations 7.10, 7.12, and 7.13) can be applied to cyclist traffic, even if it is not lane bound. Note that the distance travelled will be length, and the space–time area can be either a road length times a time, or the road area times a time. Depending on that choice, density is expressed as cyclist per metre, or cyclists per square metre. For pedestrian traffic, a density expression per square metre is most common; for cyclist traffic, both expressions are used depending on whether cyclists can move freely in the lateral direction (i.e., what type of infrastructure is considered: a narrow pathway or a large square).

On the macroscopic level, parts of the properties of bicycle flows along a road have been analysed by Botma and Papendrecht (1991) and Navin (1994). It can be assumed that bicycle traffic should obey some sort of fundamental diagram, but this has not been fully confirmed with empirical (or experimental) data. In fact, congested part of the fundamental diagram is not thoroughly established yet. Zhang et al. (2013) have made a very interesting comparison of fundamental diagrams for cars, cyclists, and pedestrians. They found the fundamental

diagrams scale by the size of the traffic participant and the speed; surprisingly and interestingly, also the states in between (combinations of density and speed) scaled similarly. Very recent research (Hogetoorn, 2022) has experimentally established the processes going on for low average speeds: in that case, the speeds do not gradually decrease to zero (cyclists cannot cycle very slowly). What happens is that a higher density leads to a higher fraction of stopped cyclists, but not necessarily a lower speed of the riding cyclists.

Queuing theory and shockwave theory would be applicable. Queuing at intersections is a more studied area of cyclist traffic. Wierbos et al. (2019) studied the capacity of cyclist traffic and how this depends on the width of the road. In short, the capacity increased more or less linearly with the road width. An interesting observation is that the capacity of the traffic leaving a queue at a traffic light is influenced by the density of the queue: the closer the people are together, the higher the flow of traffic once the traffic light turns green. This can also be used in control: if cyclists are put closer together on purpose, the outflow will increase beyond the capacity values obtained in regular traffic (Wierbos et al., 2021).

Modelling of cyclist traffic is done, often by means of the social force model (Helbing and Molnar, 1995), which is adapted for cyclist traffic (e.g., Anvari et al., 2015 for shared space). How cyclists come to a stop when approaching a traffic light is separately modelled (Gavriilidou et al., 2019). Macroscopic models for cyclist traffic are rare. These models can use a form of a fundamental diagram, which is – as mentioned – not universally accepted. A more elaborate framework, including interactions with other modes, is available (e.g., Wierbos et al., 2020), yet requires more experimental validation.

With regard to the operations on the network scale (macroscopic fundamental diagram), cyclists are smaller than cars and can cross intersections next to each other. Hence, a higher fraction of cyclists would increase the maximum flow for a network. Loder et al. (2021) use this and try to model a multi-modal macroscopic fundamental diagram.

7.12   CONCLUSIONS

Traffic flow theory entails the knowledge of the fundamental characteristics of traffic flows. Traffic flow theory and modelling is important, among others, in order to design comfortable and safe roads, to solve road congestion problems and to design adequate traffic management measures.

In traffic flow theory a basic distinction is made between microscopic and macroscopic traffic flow variables. Microscopic traffic flow variables focus on individual drivers. Macroscopic traffic flow variables reflect the average state of the traffic flow.

The fundamental diagram describes a statistical relation between the macroscopic flow variables: flow, density, and speed. The basic premise underlying the fundamental diagram is that under similar traffic conditions drivers will behave in a similar way.

Vehicle automation in its current state, without communication, does not improve road capacity. Future developments with vehicle-to-vehicle communication can do so.

Traffic flow models can be used to simulate the traffic, for instance, to evaluate *ex-ante* the use of a new part of the infrastructure. Models can be categorized based on, firstly, representa-

tion of the traffic flow in terms of flows (macroscopic), groups of drivers (macroscopic), or individual drivers (microscopic). Secondly, on underlying behavioural theory which can be based on characteristics of the flow (macroscopic) or individual drivers (microscopic behaviour).

The overall dynamics of a traffic network can be described using a remarkably simple relation, referred to as the macroscopic or network fundamental diagram (NFD). This relation shows one of the most important properties of network traffic operations, namely that its performance decreases when the number of vehicles becomes larger.

Theories and modelling tools (on a microscopic level, on a macroscopic level, and on a network level) developed for car traffic are – to a certain extent – also applicable for other modes.

REFERENCES

Anvari, B., M.G. Bell, A. Sivakumar and W.Y. Ochieng (2015), 'Modelling shared space users via rule-based social force model', *Transportation Research Part C: Emerging Technologies*, 51, 83–103.

Botma, H. (1999), 'The free speed distribution of drivers: Estimation approaches', in P. Bovy (ed.), *Five Years Crossroads of Theory and Practice*, Delft: Delft University Press, 1–22.

Botma, H. and H. Papendrecht (1991), 'Traffic Operation of Bicycle Traffic', *Transportation Research Record: Journal of the Transportation Research Board*, 1320 (1), 65–72.

Brackstone, M. and M. McDonald (1999), 'Car-Following: A Historical Review', *Transportation Research F*, 2, 181–86.

Brilon, W., R. Koenig and R.J. Troutbeck (1999), 'Useful estimation procedures for critical gaps', *Transportation Research Part A: Policy and Practice*, 33 (3–4), 161–86.

Buckley, D. (1968), 'A semi-Poisson model for traffic flow', *Transportation Science*, 2, 107–33.

Cassidy, M.J. and R.L. Bertini (1999), 'Some traffic features at freeway bottlenecks', *Transportation Research Part B: Methodological*, 33 (1), 25–42.

Chandler, R.E., R. Herman and E.W. Montroll (1958), 'Traffic dynamics: studies in car following', *Operations Research*, 6 (2), 165–84.

Chung, K., J. Rudjanakanoknada and M.J. Cassidy (2007), 'Relation between traffic density and capacity drop at three freeway bottlenecks', *Transportation Research Part B: Methodological*, 41 (1), 82–95.

Ciuffo, B., K. Mattas, A. Anesiadou and M. Makridis (2020), 'Open ACC database', accessed 23 May 2023 at http://data.europa.eu/89h/9702c950-c80f-4d2f-982f-44d06ea0009f, technical report. Brussels: European Commission, Joint Research Centre (JRC).

Chung, K., J. Rudjanakanoknada and M.J. Cassidy (2007), 'Relation between traffic density and capacity drop at three freeway bottlenecks', *Transportation Research Part B: Methodological*, 41 (1), 82–95.

Cowan, R. J. (1975), 'Useful headway models', *Transportation Research*, 9 (6), 371–75.

Daganzo, C.F. (2002a), 'A behavioral theory of multi-lane traffic flow. Part I: Long homogeneous freeway sections', *Transportation Research Part B: Methodological*, 36 (2), 131–58.

Daganzo, C.F. (2002b), 'A behavioral theory of multi-lane traffic flow. Part II: Merges and the onset of congestion', *Transportation Research Part B: Methodological*, 36 (2), 159–69.

Daganzo, C.F., V.V. Gayah and E.J. Gonzales (2011), 'Macroscopic relations of urban traffic variables: Bifurcations, multivaluedness and instability', *Transportation Research Part B: Methodological*, 45 (1), 278–88.

Daganzo, C.F. and N. Geroliminis (2008), 'An analytical approximation for the macroscopic fundamental diagram of urban traffic', *Transportation Research Part B: Methodological*, 42 (9), 771–81.

Edie, L. (1965), 'Discussion of traffic stream measurements and definitions', in J. Almond (ed.), *Proceedings of the Second International Symposium on the Theory of Road Traffic Flow*, Paris: OECD, 139–54.

Forbes, T., H. Zagorski, E. Holshouser and W. Deterline (1958), 'Measurement of driver reactions to tunnel conditions', *Highway Research Board Proceedings*, 37, 60–66.

Gavriilidou, A., W. Daamen, Y. Yuan and S.P. Hoogendoorn (2019), 'Modelling cyclist queue formation using a two-layer framework for operational cycling behaviour', *Transportation Research Part C: Emerging Technologies*, 105, 468–84.

Geroliminis, N. and C.F. Daganzo (2008), 'Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings', *Transportation Research Part B: Methodological*, 42 (9), 759–70.

Hall, F.L. and K. Agyemang-Duah (1991), 'Freeway Capacity Drop and the Definition of Capacity', *Transportation Research Record: Journal of the Transportation Research Board*, 1320, 91–98.

Helbing, D. and P. Molnar (1995), 'Social force model for pedestrian dynamics', *Physical review E*, 51 (5), 4282.

Herman, R. (1959), 'Traffic dynamics: Analysis of stability in car-following', *Operation Research*, 7 (1), 86–106.

Hogetoorn, M. (2022), 'Dense cycling conditions: The influence of stopped cyclists on the flow of bicycle traffic'. MSc thesis, Delft University of Technology.

Hoogendoorn, S.P. (2005), 'Unified approach to estimating free speed distributions', *Transportation Research Part B: Methodological*, 39 (8), 709–27.

Hoogendoorn, R., S.P. Hoogendoorn, K. Brookhuis and W. Daamen (2010), 'Mental workload, longitudinal driving behavior, and adequacy of car-following models for incidents in other driving lane', *Transportation Research Record*, 2188 (1), 64–73.

Jepsen, M. (1998), 'On the speed-flow relationships in road traffic: A model of driver behaviour', *Proceedings of the Third International Symposium on Highway Capacity*, Copenhagen, 297–319.

Kerner, B.S. (2004), *The Physics of Traffic: Empirical Freeway Pattern Features, Engineering Applications, and Theory*. Berlin: Springer.

Kesting, A., M. Treiber and D. Helbing (2007), 'General lane-changing model MOBIL for car-following models', *Transportation Research Record*, 1999 (1), 86–94.

Keyvan-Ekbatani, M., X.(S.) Gao, V.V. Gayah and V.L. Knoop (2019), 'Traffic-responsive signals combined with perimeter control: investigating the benefits', *Transportmetrica B: Transport Dynamics*, 7 (1), 1402–25.

Keyvan-Ekbatani, M., A. Kouvelas, I. Papamichail and M. Papageorgiou (2012), 'Exploiting the fundamental diagram of urban networks for feedback-based gating', *Transportation Research Part B: Methodological*, 46 (10), 1393–403.

Knoop, V.L. (2020), *Traffic Flow Theory: An Introduction with Exercises. Third edition*, Delft: TU Delft Open Textbook. DOI: 10.5074/t.2021.002.

Knoop, V.L., A. Duret, C. Buisson and B. Van Arem (2010, September), 'Lane distribution of traffic near merging zones influence of variable speed limits', in *13th International IEEE Conference on Intelligent Transportation Systems*, Madeira: IEEE, 485–90.

Knoop, V.L. and S.P. Hoogendoorn (2022), 'Free Flow Capacity and Queue Discharge Rate: Long Term Changes', *Transportation Research Records*. DOI: 10.1177/03611981221078845.

Knoop, V.L., S.P. Hoogendoorn and H.J. Van Zuylen (2009), 'Empirical Differences between Time Mean Speed and Space Mean Speed', in C. Appert-Rolland, F. Chevoir, P. Gondret, S. Lassarre, J.-P. Lebacque and M. Schreckenberg (ed.), *Proceedings of Traffic and Granular Flow 07*, New York: Springer, 351–56.

Knoop, V.L., M. Wang, I.M. Wilmink, M. Hoedemaeker, M. Maaskant and E.-J. Van der Meer (2019), 'Platoon of SAE level-2 automated vehicles on public roads: setup, traffic interactions, and stability', *Transportation Research Records*, 2673 (9), 311–22.

Laval, J.A. (2011), 'Hysteresis in traffic flow revisited: An improved measurement method', *Transportation Research Part B: Methodological*, 45 (2), 385–91.

Laval, J.A. and C.F. Daganzo (2006), 'Lane-changing in traffic streams', *Transportation Research Part B: Methodological*, 40 (3), 251–64.

Leutzbach, W. and R. Wiedemann (1986), 'Development and applications of traffic simulation models at the Karlsruhe Institut für Verkehrswesen', *Traffic Engineering and Control*, 27, 270–78.

Loder, A., L. Ambühl, M. Menendez and K.W. Axhausen (2019), 'Understanding traffic capacity of urban networks', *Scientific reports*, 9 (1), 1–10.

Loder, A., L. Bressan, M.J. Wierbos, H. Becker, H. Emmonds, M. Obee, V.L. Knoop, M. Menendez and K.W. Axhausen (2021), 'How many cars in the city are too many? Towards finding the optimal modal split for a multi-modal urban road network', *Frontiers in Future Transportation*, 2, 665006.

May, A.D. (1990), *Traffic Flow Fundamentals*, Englewood Cliffs, NJ: Prentice Hall.

Michaels, R. (1963), 'Perceptual factors in car following', , in J. Almond (ed.), *Proceedings of the Second International Symposium on the Theory of Road Traffic Flow*, Paris: OECD, 44–59.

Milanés, V. and S.E. Shladover (2014), 'Modeling cooperative and autonomous adaptive cruise control dynamic responses using experimental data', *Transportation Research Part C: Emerging Technologies*, 48, 285–300.

Minderhoud, M.M., H. Botma and P.H.L. Bovy (1996), 'An Assessment of Roadway Capacity Estimation Methods', Technical Report vk2201.302, Delft University of Technology.

Nagalur Subraveti, H.H.S., V.L. Knoop and B. Van Arem (2020), 'Improving Traffic Flow Efficiency at Motorway Lane Drops by Influencing Lateral Flows', *Transportation Research Records*, 2674 (11) 367–78.

National Academies of Sciences, Engineering, and Medicine (2022), *Highway Capacity Manual 7th Edition: A Guide for Multimodal Mobility Analysis*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26432.

Navin, F.P.D. (1994), 'Bicycle Traffic Flow Characteristics: Experimental Results and Comparisons', *ITE Journal*, 64 (3), 31–37.

Nelson, W. (1982), *Applied Life Time Analysis*, New York: Wiley.

Newell, G.F. (1961), 'A theory of traffic flow in tunnels', in R. Herman (ed.), *Theory of Traffic Flow, Proceedings of the Symposium on the Theory of Traffic flow*, Amsterdam: Elsevier, 193–206.

Pignataro, L. (1973), *Traffic Engineering: Theory and Practice*, Englewood Cliffs, NJ: Prentice Hall.

Pipes, L. (1953), 'Car following models and the fundamental diagram of road traffic', *Transportation Research*, 1, 21–29.

Pueboobpaphan, R. and B. van Arem (2010, January). 'Understanding the relation between driver/ vehicle characteristics and platoon/traffic flow stability for the design and assessment of cooperative adaptive cruise control', [paper 10–0994 on DVD], in *89th Transportation Research Board (TRB) Annual Meeting 2010*, Washington, DC: Mira Digital Publishing.

Rao, B., P. Varaiya and F. Eskafi (1993), 'Investigations into achievable capacities and stream stability with coordinated intelligent vehicles', *Transportation Research Record: Journal of the Transportation Research Board*, 1408, 27–35.

Roncoli, C., N. Bekiaris-Liberis and M. Papageorgiou (2017), 'Lane-changing feedback control for efficient lane assignment at motorway bottlenecks', *Transportation Research Record*, 2625 (1), 20–31.

SAE, J3016 (2021), 'Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems'.

Schakel, W.J., V.L. Knoop and B. Van Arem (2012), 'LMRS: Integrated Lane Change Model with Relaxation and Synchronization', *Transportation Research Records: Journal of the Transportation Research Board*, 2316, 47–57.

Shiomi, Y., J. Xing, H. Kai, and T. Katayama (2019), 'Analysis of the Long-Term Variations in Traffic Capacity at Freeway Transportation Research Record Bottleneck', *Transportation Research Record: Journal of the Transportation Research Board*, 2673, 390–401.

Thomson, J.M. (1967), 'Speeds and flows of traffic in Central London: Speed-flow relations', *Traffic Engineering and Control*, 8 (12), 721–25.

Treiber, M., A. Hennecke and D. Helbing (2000), 'Congested traffic states in empirical observations and microscopic simulations', *Physical review E*, 62 (2), 1805.

Treiterer, J. and J.A. Myers (1974), 'The hysteresis phenomenon in traffic flow', in D.J. Buckley (ed.), *Proceedings of the 6th International Symposium on Transportation and Traffic Theory*, New York: Elsevier, 13–38.

Van Erp, P.B., V.L. Knoop and S.P. Hoogendoorn (2019), 'On the value of relative flow data', *Transportation Research Part C: Emerging Technologies*, 113, 74–90.

Van Wageningen-Kessels, F., H. Van Lint, K. Vuik and S. Hoogendoorn (2015), 'Genealogy of traffic flow models', *EURO Journal on Transportation and Logistics*, 4 (4), 445–73.

Wardrop, J.G. (1968), 'Journey speed and flow in central urban areas', *Traffic Engineering and Control*, 9 (11), 528–32.

Wasielewski, P. (1974), 'An integral equation for the semi-Poisson headway distribution', *Transportation Science*, 8, 237–47.

Wierbos, M.J., V.L. Knoop, R.L. Bertini and S.P. Hoogendoorn (2021), 'Influencing the queue config-uration to increase bicycle jam density and discharge rate: An experimental study on a single path', *Transportation Research Part C: Emerging Technologies*, 122, 102884.

Wierbos, M.J., V.L. Knoop, F.S. Hänseler and S.P. Hoogendoorn (2019), 'Capacity, capacity drop, and relation of capacity to the path width in bicycle traffic', *Transportation research record*, 2673 (5), 693–702.

Wierbos M.J., V.L. Knoop, F.S. Hänseler and S.P. Hoogendoorn (2020), 'A Macroscopic Flow Model for Mixed Bicycle–Car Traffic', *Transportmetrica A: Transport Science*, 17(3), 9935.

Yuan, K., V.L. Knoop and S.P. Hoogendoorn (2015), 'Capacity drop: a relation between the speed in congestion and the queue discharge rate', in *Transportation Research Record*, 2491, 72–80.

Yuan, K., V.L. Knoop and S.P. Hoogendoorn (2017), 'A microscopic investigation into the capacity drop: impacts of longitudinal behavior on the queue discharge rate', *Transportation Science*, 51 (3), 852–62.

Zahavi, Y. (1972), 'Traffic performance evaluation of road networks by the α-relationship', Parts I and II, *Traffic Engineering and Control*, 14 (5 and 6), 228–31 and 292–93.

Zhang, J., W. Mehner, E. Andresen, S. Holl, M. Boltes, A. Schadschneider and A. Seyfried (2013), 'Comparative Analysis of Pedestrian, Bicycle and Car Traffic Moving in Circuits', *Procedia–Social and Behavioral Sciences*, 104, 1130–38.