Introduction to Modern Traffic Flow Theory and Control

Boris S. Kerner

Introduction to Modern Traffic Flow Theory and Control

The Long Road to Three-Phase Traffic Theory

Springer

Boris S. Kerner
Daimler AG
GR/PTI
HPC: G021
71059 Sindelfingen
Germany
boris.kerner@daimler.com

ISBN 978-3-642-02604-1 e-ISBN 978-3-642-02605-8
DOI 10.1007/978-3-642-02605-8
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2009933980

Cover image: © WONG SZE FEI - Fotolia.com

Cover design: deblik, Berlin

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The understanding of empirical traffic congestion occurring on unsignalized multi-lane highways and freeways is a key for effective traffic management, control, organization, and other applications of transportation engineering. However, the traffic flow theories and models that dominate up to now in transportation research journals and teaching programs of most universities cannot explain either traffic breakdown or most features of the resulting congested patterns. These theories are also the basis of most dynamic traffic assignment models and freeway traffic control methods, which therefore are not consistent with features of real traffic.

For this reason, the author introduced an alternative traffic flow theory called three-phase traffic theory, which can predict and explain the empirical spatiotemporal features of traffic breakdown and the resulting traffic congestion. A previous book "The Physics of Traffic" (Springer, Berlin, 2004) presented a discussion of the empirical spatiotemporal features of congested traffic patterns and of three-phase traffic theory as well as their engineering applications.

Rather than a comprehensive analysis of empirical and theoretical results in the field, the present book includes no more empirical and theoretical results than are necessary for the understanding of vehicular traffic on unsignalized multi-lane roads. The main objectives of the book are to present an "elementary" traffic flow theory and control methods as well as to show links between three-phase traffic theory and earlier traffic flow theories. The need for such a book follows from many comments of colleagues made after publication of the book "The Physics of Traffic".

Another important objective of this book is to give an introduction to methods of spatiotemporal traffic congestion recognition and prediction, on-ramp metering, speed limit control, and some other freeway control and dynamic management methods whose theoretical basis is three-phase traffic theory. The importance of this subject can be explained as follows. Almost all other traffic flow theories and the associated freeway control and dynamic management methods assume the existence of a *particular* (fixed or stochastic) highway capacity of free flow at a highway bottleneck and, therefore, they use the highway capacity as a basic parameter of dynamic traffic management models. In this book we show and explain how and why the application of a particular highway capacity in methods for dynamic freeway traffic management like on-ramp metering, speed limit control, or dynamic traffic assignment, is not consistent with features of real traffic.

Through an application of the principle "no more results than are necessary", I hope to present traffic flow theory and control in a manner understandable to a broad audience of readers interested in traffic phenomena. With this aim, the book also includes an extended glossary with definitions and explanations of terms used.

I thank Ralf G Herrtwich and Matthias Schulze for their support as well as many other my colleagues at the Daimler Company, in particular, Hubert Rehborn, Gerhard Nöcker, Andreas Hiller, Achim Brakemaier, Ines Maiwald-Hiller, Winfried Kronjäger for fruitful discussions and advice. I thank also Dietrich Wolf for useful suggestions. Particular thanks are to Achim Brakemaier, Viktor Friesen, Sergey Klenov, Gerhard Nöcker, Andreas Hiller, Winfried Kronjäger, Jochem Palmer, and Hubert Rehborn who have read the book and made many useful comments. I thank also Hesham Rakha, Hani Mahmassani, and Jorge Laval for helpful discussions about approaches to traffic flow modeling in Woods Hole in July 2008. Many thanks to Rüdiger Hain, Oliver Baumann and all other friends who have encouraged me while writing this book. I am grateful to Sergey Klenov for his help with numerical simulations and the preparation of illustrations for the book. Finally, I thank my wife, Tatiana Kerner, for her help and understanding.

Stuttgart, May 2009 *Boris Kerner*

Contents

Acronyms and Symbols

F Free traffic flow
C Congested traffic
S Synchronized flow phase of congested traffic
J Wide moving jam phase of congested traffic
Line *J* Characteristic line in the flow–density plane representing a steadily propagation of the downstream front of a wide moving jam. The slope of the line *J* is determined by the mean velocity of the downstream jam front
F→S transition Traffic breakdown, i.e., phase transition from the free flow phase to the synchronized flow phase
S→J transition Phase transition from the synchronized flow phase to the wide moving jam phase
F→S→J transitions Sequence of an F→S transition with a following S→J transition
SP Synchronized flow pattern
LSP Localized SP
WSP Widening SP
MSP Moving SP
GP General congested traffic pattern
EP Expanded traffic congested pattern
ESP Expanded synchronized flow pattern
EGP Expanded general pattern
FOTO **F**ocasting **O**f **T**raffic **O**bjects is a macroscopic model for automatic traffic phase reconstruction and tracking of synchronized flow
ASDA **A**utomatische **S**tau**d**ynamik **A**nalyse (automatic tracking of moving traffic jams) is a macroscopic model for tracking of moving jams
ANCONA **A**utomatic o**n**-ramp **c**ontrol of c**on**gested p**a**tterns is a control approach in which congestion is allowed to set in at a bottleneck. The basic idea is to maintain congestion conditions at the bottleneck to the minimum possible level; in particular, a congested pattern should not propagate upstream
ACC Adaptive cruise control

v vehicle speed [km/h] or [m/s]
v_ℓ speed of the preceding vehicle [km/h] or [m/s]
$\Delta v = v_\ell - v$ speed difference between the speed of the preceding vehicle and the vehicle speed [km/h] or [m/s]
a vehicle acceleration (deceleration) $[\mathrm{m/s}^2]$
q flow rate [vehicles/h] or [vehicles/min]
ρ vehicle density [vehicles/km]
g space gap between vehicles [m] that is also called as net distance or space headway
τ time headway between vehicles [s] that is also called as time gap or net time distance
d vehicle length [m]
o occupancy [%]
v_{g} The mean velocity of the downstream front of a wide moving jam
q_{out} The flow rate in free flow formed in the outflow of a wide moving jam
ρ_{min} The vehicle density in free flow formed in the outflow of a wide moving jam
$q_{\mathrm{th}}^{(\mathrm{B})}$ Threshold flow rate for traffic breakdown at a bottleneck that is the minimum highway capacity of free flow
$q_{\mathrm{max}}^{(\mathrm{free\ B})}$ Maximum flow rate in free flow downstream of a bottleneck that is the maximum highway capacity of free flow
$q^{(\mathrm{cong})}$ The average flow rate within a congested pattern upstream of a bottleneck averaged during a time interval that is considerably longer than time distances between any moving jams within the congested pattern
q_{in} The flow rate in free flow on the main road upstream of a bottleneck
q_{on} The on-ramp inflow rate at an on-ramp bottleneck
q_{sum} The flow rate in free flow downstream of a bottleneck under condition that free flow is at the bottleneck
q_{off} The flow rate of vehicles leaving the main road to off-ramp at an off-ramp bottleneck
$v_{\mathrm{cr,\ FS}}^{(\mathrm{B})}$ A critical speed required for traffic breakdown at a bottleneck
$\rho_{\mathrm{cr,\ FS}}^{(\mathrm{B})}$ A critical density required for traffic breakdown at a bottleneck
$v_{\mathrm{cr}}^{(\mathrm{SJ})}$ A critical speed in metastable synchronized flow required for wide moving jam emergence
$\rho_{\mathrm{cr}}^{(\mathrm{SJ})}$ A critical density in metastable synchronized flow required for wide moving jam emergence
$\tau_{\mathrm{del}}^{(\mathrm{a})}$ The mean time delay in vehicle acceleration
$\tau_{\mathrm{del,\ jam}}^{(\mathrm{a})}$ The mean time delay in vehicle acceleration at the downstream front of a wide moving jam
g_{safe} A safe space gap between vehicles
τ_{safe} A safe time headway (safe time gap) between vehicles
G A synchronization space gap between vehicles

τ_{G}	A synchronization time headway (synchronization time gap) between vehicles
τ_{reac}	The mean driver reaction time
T_{av}	Averaging time interval for traffic variables
$P_{\mathrm{FS}}^{(\mathrm{B})}$	Probability of traffic breakdown at a bottleneck
$P_{\mathrm{C}}^{(\mathrm{B})}$	Probability that free flow remains at a bottleneck during time interval T_{av}
τ_{step}	A time step in traffic flow models with a discrete time
$r = \mathrm{rand}(0,1)$	A random number uniformly distributed between 0 and 1

Chapter 1
Introduction

Vehicular traffic is an extremely complex dynamic process associated with the spatiotemporal behavior of many-particle systems. The complexity of vehicular traffic is due to nonlinear interactions between the following three main dynamic processes (Fig. 1.1):

(i) travel *decision* behavior, which determines traffic demand,
(ii) *routing* of vehicles in a traffic network, and
(iii) *traffic congestion* occurrence within the network.

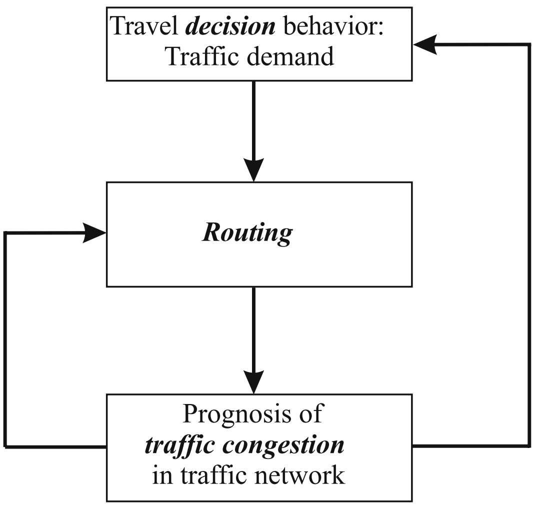

Fig. 1.1 Explanation of complexity of vehicular traffic

Travel decision behavior determines travel demand. Traffic routing in the network is associated with traffic supply. However, traffic congestion occurring within

B.S. Kerner, *Introduction to Modern Traffic Flow Theory and Control*,
DOI 10.1007/978-3-642-02605-8_1, © Springer-Verlag Berlin Heidelberg 2009

the traffic network restricts free flow travel. This influences both travel decision behavior and traffic routing in the network. Indeed, because of traffic congestion, a person decides to stay at home or travel by train rather than by car. A feedback between *traffic congestion* and *travel decision* is symbolically shown by arrow on the right hand side in Fig. 1.1. In turn, because of traffic congestion on a route from an origin to a destination usually used, a person changes the route of travel. A feedback between *traffic congestion* and *routing* is symbolically shown by arrow on the left hand side in Fig. 1.1.

Empirical traffic congestion, i.e., traffic congestion observed in real measured traffic data is a *spatiotemporal effect*: The traffic congestion occurs in space and time in the form of spatiotemporal congested traffic patterns propagating within a traffic network. These empirical congested traffic patterns exhibit a variety of complex spatiotemporal features. For this reason, the complexity of traffic management is associated with this variety of the empirical congested traffic patterns as well as with the necessity in the optimization of these patterns. This optimization should ensure either the dissolution of traffic congestion or, if this is not possible to achieve, the minimization of the influence of traffic congestion on travel costs.

- We see that the understanding of *empirical traffic congestion* is the key for effective traffic management, control, organization, and all other applications of transportation engineering.

Inputs to travel decision behavior models are the typical regional model data about social, economic, and demographic information of potential travelers and land use information to create schedules followed by people in their everyday life. The output are detailed lists of activities pursued, times spent in each activity, and travel information from activity to activity. A review of travel decision behavior models has recently been done by Goulias [1].

Routing based on a traffic optimization, which is associated with a minimization of chosen travel "costs", together with a prognosis of traffic congestion is called *dynamic traffic assignment* in the traffic network. A dynamic traffic assignment model should find the link inflows for the traffic network. The model includes usually a traffic flow model, which makes a prognosis of traffic in the network, and a routing model associated with the problem of traffic optimization. The router model computes the sequence of roadways that minimize travel costs of the traffic network. Examples of the travel costs are travel time, fuel consumption, or HC and CO_2 emissions. The traffic flow and router models are connected by a feedback loop (see the feedback loop between *traffic congestion* and *routing* in Fig. 1.1). As a result, traffic congestion in the network predicted by the traffic flow model changes results of dynamic traffic assignment considerably. For this reason, the traffic flow model should model traffic congestion as close as possible to real traffic congestion found in empirical observations. A review of models for dynamic traffic assignment in traffic networks has recently been done by Rakha and Tawfik [2]. Approaches to traffic prognosis have recently been reviewed by Rehborn and Klenov [3].

A complex spatiotemporal behavior of empirical traffic congested patterns was studied during the last 75 years by several generations of researchers (see references in reviews and books [4–26]). It was found that traffic congestion in the traffic network results from traffic breakdown in an initially free flow: vehicle speeds decrease abruptly to lower speeds in congested traffic [4, 5, 27–29]. Traffic breakdown is observed mostly at highway bottlenecks. A bottleneck can be a result of road works, on- and off-ramps, a decrease in the number of road lanes, road curves and road gradients, bad weather conditions, accidents, etc. [4, 5, 27–29]. Beginning from the classic work of Greenshields [30], the most of the traffic flow theories and models have been made in the framework of the so-called *fundamental diagram of traffic flow*. The fundamental diagram is a flow rate–density relationship, i.e., a correspondence between a given vehicle density and the flow rate in traffic flow. The fundamental diagram reflects the obvious result of empirical observations that the greater the density, the lower the averaged speed in vehicular traffic.

However, the puzzle of empirical spatiotemporal features of traffic congestion has been solved only recently [25]. As will be discussed in this book, it turns out that earlier traffic flow theories and models reviewed in [4–24] cannot explain either traffic breakdown or most features of the resulting spatiotemporal congested patterns. These traffic flow theories and models, which dominate up to now in transportation research journals and teaching programs of most universities, are also the theoretical basis for dynamic traffic assignment models and methods for freeway traffic control. Therefore, the associated methods for dynamic traffic management are not consistent with features of real traffic.

To explain empirical spatiotemporal features of vehicular traffic, the author introduced an alternative traffic flow theory called three-phase traffic theory. A consideration of empirical spatiotemporal features of congested traffic patterns and three-phase traffic theory that explains these pattern features as well as some resulting engineering applications have been presented in the previous book [25].

Rather than a comprehensive discussion congested traffic patterns, in the present book the author gives only an introduction to traffic flow theory and control on multi-lane roads[1] including no more empirical and theoretical results than are necessary for the understanding of vehicular traffic as well as to make a more detailed consideration of links between three-phase traffic theory and earlier traffic flow theories. The main objectives of this book are as follows:

(1) To explain why rather than the fundamental diagram of traffic flow, *spatiotemporal* analysis of empirical congested traffic patterns is the key for the understanding of traffic flow characteristics as well as for the development of dynamic traffic management methods (including methods for dynamic traffic control and assignment) that are consistent with real traffic.

[1] In the book, we limit attention to dynamic traffic phenomena determined by driver interactions in traffic. These traffic phenomena play the most important role on freeways and highways. In contrast, in city traffic, light signals and other traffic regulations at road intersections can often almost fully determine traffic dynamics. A review about urban traffic control has recently been done by Gartner and Stamatiadis [31]. See also the UTA model for the **u**rban **t**raffic **a**nalysis and prognosis in Sect. 22.4 of the book [25].

(2) To explain why classic traffic flow theories and models cannot explain either traffic breakdown or most features of the resulting spatiotemporal congested patterns.
(3) To give a new basis for the development of models for dynamic traffic operation methods, dynamic traffic assignment models, and highway traffic control methods, which are consistent with features of real traffic.

The importance of these objectives can be explained as follows. Most earlier traffic flow theories and the associated freeway control and dynamic management methods assume the existence of a *particular* fixed or stochastic highway capacity of free flow at a highway bottleneck. Therefore, they use the highway capacity as a basic parameter of dynamic traffic management models. In this book we show and explain how and why the application of a particular highway capacity as a control parameter in methods for dynamic freeway traffic management like on-ramp metering, speed limit control, or dynamic traffic assignment, is not consistent with features of real traffic.

The book consists of two parts. Part I is devoted to a consideration of empirical spatiotemporal features and characteristics of traffic and three-phase traffic theory that explains these traffic features and characteristics. In Part II, we discuss the impact of three-phase traffic theory on traffic control and management. Because simulations of the prognosis of traffic congestion with mathematical traffic flow models can be considered a part of traffic control and management models (see Fig. 1.1 and related explanations made above), a critical discussion of the impact of three-phase traffic theory on these models has also been included in Part II.

Part I begins with traffic phase definitions made in three-phase traffic theory (Chap. 2). Explanations of empirical spatiotemporal traffic flow characteristics with three-phase traffic theory are the subject of the subsequent Chaps. 3–7.

In particular, in Chaps. 3 and 4 we explain why the fundamental empirical features of traffic breakdown at a bottleneck leads to the conclusion of three-phase traffic theory that rather than a particular highway capacity, there are the infinite number of highway capacities at the bottleneck. A consideration of the spontaneous emergence of wide moving jams is the subject of Chap. 5. The origin of some of the hypotheses and terms of three-phase traffic theory used in previous chapters of the book is discussed in Chap. 6.

In Chap. 7, we discuss a variety of spatiotemporal congested patterns arising from traffic breakdown and wide moving jam emergence.

Part II begins from a compendium of three-phase traffic theory (Chap. 8). In Chap. 9, we briefly discuss methods for spatiotemporal congested pattern reconstruction, tracking, and control.

Earlier theoretical basis of transportation engineering is discussed in Chap. 10. Here we discuss both the achievements and drawbacks of earlier traffic flow theories in explanations of real measured spatiotemporal features of traffic congestion. In particular, the critical part of this consideration contains the following subjects:

(i) That and why many of the fundamental empirical spatiotemporal features of traffic patterns are *lost* in the fundamental diagram of traffic flow.

(ii) Why the earlier traffic flow theories and models in the framework of the fundamental diagram of traffic flow, which are up to now the basic approaches in transportation research [4–24], have failed to show empirical features of traffic breakdown.

(iii) That and why well-accepted definitions of highway capacity as a particular (either fixed or stochastic) value are also not consistent with the fundamental empirical features of traffic breakdown and, as a result, methods for traffic flow control, dynamic traffic assignment as well as other methods of dynamic traffic management, which are based on these capacity definitions, are not consistent with real measured spatiotemporal traffic flow characteristics.

In Chap. 11 we discuss some mathematical traffic flow models in the framework of three-phase traffic theory. These models can simulate traffic breakdown and all resulting spatiotemporal traffic flow characteristics as they are observed in real measured traffic data.

A discussion of links between three-phase traffic theory and the fundamental diagram approach to traffic flow modeling is the subject of Chap. 12. In this discussion we would like to answer the question what features of three-phase traffic theory are missing in the earlier traffic flow theories of Chap. 10 resulting in the failure of these theories in the explanation of traffic breakdown as observed in real measured traffic data.

References

1. K.G. Goulias, in *Encyclopedia of Complexity and System Science*, ed. by R.A. Meyers. (Springer, Berlin, 2009), pp. 9536–9565
2. H. Rakha, A. Tawfik, in *Encyclopedia of Complexity and System Science*, ed. by R.A. Meyers. (Springer, Berlin, 2009), pp. 9429–9470
3. H. Rehborn, S.L. Klenov, in *Encyclopedia of Complexity and System Science*, ed. by R.A. Meyers. (Springer, Berlin, 2009), pp. 9500–9536
4. A.D. May, *Traffic Flow Fundamentals*, (Prentice-Hall, Inc., New Jersey, 1990)
5. *Highway Capacity Manual 2000*, (National Research Council, Transportation Research Boad, Washington, D.C., 2000)
6. F.A. Haight, *Mathematical Theories of Traffic Flow*, (Academic Press, New York, 1963)
7. I. Prigogine, R. Herman, *Kinetic Theory of Vehicular Traffic*, (American Elsevier, New York, 1971)
8. W. Leutzbach, *Introduction to the Theory of Traffic Flow*, (Springer, Berlin, 1988)
9. N.H. Gartner, C.J. Messer, A.K. Rathi (eds), *Traffic Flow Theory: A State-of-the-Art Report*, (Transportation Research Board, Washington DC, 2001)
10. M. Cremer, *Der Verkehrsfluss auf Schnellstrassen*, (Springer, Berlin, 1979)
11. G.B. Whitham, *Linear and Nonlinear Waves*, (Wiley, New York, 1974)
12. R. Wiedemann, *Simulation des Verkehrsflusses*, (University of Karlsruhe, Karlsruhe, 1974)
13. G.F. Newell, *Applications of Queuing Theory*, (Chapman Hall, London, 1982)
14. C.F. Daganzo, *Fundamentals of Transportation and Traffic Operations*, (Elsevier Science Inc., New York, 1997)
15. M. Papageorgiou, *Application of Automatic Control Concepts in Traffic Flow Modeling and Control*, (Springer, Berlin, New York, 1983)
16. M. Brackstone, M. McDonald, Transportation Research F **2**, 181 (1998)

17. D.C. Gazis, *Traffic Theory*, (Springer, Berlin, 2002)
18. N. Bellomo, V. Coscia, M. Delitala, Math. Mod. Meth. App. Sc. **12**, 1801–1843 (2002)
19. D. Chowdhury, L. Santen, A. Schadschneider, Physics Reports **329**, 199 (2000)
20. D. Helbing, Rev. Mod. Phys. **73**, 1067–1141 (2001)
21. R. Mahnke, J. Kaupužs, I. Lubashevsky, Phys. Rep. **408**, 1–130 (2005)
22. T. Nagatani, Rep. Prog. Phys. **65**, 1331–1386 (2002)
23. K. Nagel, P. Wagner, R. Woesler, Oper. Res. **51**, 681–716 (2003)
24. B. Piccoli, A. Tosin, in *Encyclopedia of Complexity and System Science*, ed. by R.A. Meyers. (Springer, Berlin, 2009), pp. 9727–9749
25. B.S. Kerner, *The Physics of Traffic*, (Springer, Berlin, New York, 2004); in *Encyclopedia of Complexity and System Science*, ed. by R.A. Meyers. (Springer, Berlin, 2009), pp. 9302–9355; 9355–9411
26. S. Maerivoet, B. De Moor, Phys. Rep. **419**, 1-64 (2005)
27. F.L. Hall, K. Agyemang-Duah, Trans. Res. Rec. **1320**, 91–98 (1991)
28. L. Elefteriadou, R.P. Roess, W.R. McShane, Transp. Res. Rec. **1484**, 80–89 (1995)
29. B.N. Persaud, S. Yagar, R. Brownlee, Trans. Res. Rec. **1634**, 64–69 (1998)
30. B.D. Greenshields, in *Highway Research Board Proceedings*, **14**, pp. 448-477 (1935)
31. N.H. Gartner, Ch. Stamatiadis, in *Encyclopedia of Complexity and System Science*, ed. by R.A. Meyers. (Springer, Berlin, 2009), pp. 9470–9500

Part I
Three-Phase Traffic Theory

Chapter 2
Definitions of The Three Traffic Phases

2.1 Traffic Variables, Parameters, and Patterns

Traffic flow phenomena are associated with a complex dynamic behavior of spatiotemporal traffic patterns. The term *spatiotemporal* reflects the empirical evidence that traffic occurs in *space and time*. Therefore, only through a *spatiotemporal* analysis of real measured traffic data the understanding of features of real traffic is possible. In other words, spatiotemporal features of traffic can only be found, if traffic variables are measured in real traffic in space and time.

The term a *spatiotemporal traffic pattern* (traffic pattern for short) is defined as follows:

- A spatiotemporal traffic pattern is a distribution of traffic flow variables in space and time.

Examples of traffic variables are the flow rate q [vehicles/h], vehicle density ρ [vehicles/km], and vehicle speed v [km/h] or [m/s] (see, e.g., [1–3]).

The term *empirical* features of a spatiotemporal traffic pattern means that the features are found based on an analysis of real measured traffic data.

A spatiotemporal traffic pattern is limited spatially by pattern fronts. There are downstream and upstream fronts of a traffic pattern. The downstream pattern front separates the pattern from other traffic patterns downstream. The upstream pattern front separates the pattern from other traffic patterns upstream.

The term *front of traffic pattern* is defined as follows:

- A front of a traffic pattern is either a moving or motionless region within which one or several of the traffic variables change abruptly in space (and in time, when the front is a moving one).

Traffic variables and traffic patterns can depend considerably on *traffic parameters*.

- Traffic parameters are parameters, which can influence traffic variables and traffic patterns.

B.S. Kerner, *Introduction to Modern Traffic Flow Theory and Control*,
DOI 10.1007/978-3-642-02605-8_2,

Examples of traffic parameters are a traffic network infrastructure (including, e.g., highway bottleneck types and their locations), weather (whether the day is sunny or rainy or else foggy, dry or wet road, or even ice and snow on road), percentage of long vehicles, day time, working day or week-end, other road conditions, and vehicle technology.

Considering traffic flow patterns, we distinguish between *macroscopic* and *microscopic* descriptions of the patterns.

In the macroscopic pattern description, the behavior of macroscopic measured traffic variables and macroscopic characteristics of traffic flow patterns in space and time should be studied and understood.

Examples of the macroscopic traffic variables are the flow rate, vehicle density, occupancy, and average vehicle speed (see, e.g., [1–3]).

An example of macroscopic characteristics of a traffic pattern is the mean velocities of the downstream and upstream fronts of the pattern. We see that the macroscopic traffic variables and pattern characteristics are associated with an averaging behavior of many vehicles in traffic, i.e., the variables and characteristics are averaged during an averaging time interval for traffic variables denoted by $T_{\rm av}$.

As an example of the term an *averaging time interval for traffic variables*, we consider *1-min average data* that means the following: all macroscopic traffic variables associated with a traffic pattern under consideration are averaged with the use of the same averaging time interval $T_{\rm av} = 1$ min.

In contrast with the macroscopic description of traffic patterns, the microscopic description of traffic flow patterns is associated with a study of microscopic traffic variables and microscopic pattern characteristics that reflect the behavior of individual (called also *single*) vehicles in traffic flow.

Examples of the microscopic traffic flow variables are single vehicle space coordinates and their time-dependence, a time headway (net time distance) τ [s] and a space gap (net distance) g [m] between two vehicles following each other (Fig. 2.1), a single vehicle speed v [km/h] or [m/s], a vehicle length d [m] [1–3]. In particular, vehicle space coordinates and their time-dependence can be used for the reconstruction of vehicle trajectories, i.e., the trajectories of vehicles in the space–time plane[1]. Note that measured traffic data in which microscopic traffic variables can be identified are also called *single vehicle data*[2].

There are many measurement techniques of traffic flow variables based on road detectors (see, e.g., [1–3]), video camera measurements (see, e.g., [4]), etc. We briefly discuss measurements of traffic variables with induction double loop detectors installed at some road locations.

Each detector consists of two induction loops spatially separated by a given small distance $\ell_{\rm d}$ (Fig. 2.1). The induction loop registers a vehicle moving on the road by producing a pulse electric current that begins at some time $t_{\rm b}$ when the vehicle reaches the induction loop and it ends some time later $t_{\rm f}$ when the vehicle leaves the

---

[1] An example of empirical vehicle trajectories is shown in Fig. 2.3 of Sect. 2.2.1.

[2] Naturally, there is also an intermediate description of traffic called as a *mesoscopic* description of traffic phenomena in which both macroscopic and microscopic traffic flow variables and/or characteristics of traffic patterns are studied.

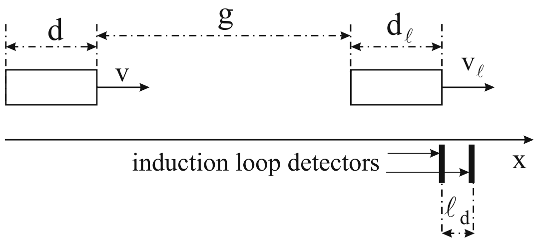

Fig. 2.1 Qualitative scheme of induction loop detector measurements

induction loop. The duration of this current pulse

$$\Delta t = t_{\mathrm{f}} - t_{\mathrm{b}} \tag{2.1}$$

is therefore related to the time taken by the vehicle to traverse the induction loop.

Every vehicle that passes the induction loop produces a related current pulse. This enables us to calculate the gross time gap between the vehicle with a speed v and the preceding vehicle with a speed v_ℓ that have passed the induction loop one after the other:

$$\tau^{(\mathrm{gross})} = t_{\mathrm{b}} - t_{\ell,\,\mathrm{b}}, \tag{2.2}$$

where subscript ℓ is related to the preceding vehicle (Fig. 2.1). We can further calculate the flow rate q as the measured number of vehicles N passing the induction loop during a given averaging time interval for traffic variables T_{av}:

$$q = \frac{N}{T_{\mathrm{av}}}. \tag{2.3}$$

Because there are two different induction loops in each detector, separated by a known distance ℓ_{d} from one another, the detector is able to measure the individual vehicle speed. Indeed, due to the distance ℓ_{d} between two loops of the detector, the first (upstream) loop registers the vehicle earlier than the second (downstream) one. Therefore, if the vehicle speed v is not zero, there will be a time lag δt between the current pulses produced by the two detector induction loops when the vehicle passes both. It is assumed that by virtue of the small value of ℓ_{d}, the vehicle speed does not change between the induction loops. This enables us to calculate the single (individual) vehicle speed v:

$$v = \frac{\ell_{\mathrm{d}}}{\delta t} \tag{2.4}$$

and the vehicle length d

$$d = v\Delta t. \tag{2.5}$$

From Eqs. (2.2) and (2.5) it is possible to calculate the time headway:

$$\tau = \tau^{(\mathrm{gross})} - \frac{d_\ell}{v_\ell}. \tag{2.6}$$

At a given time instant $t = t_1$, the time headway between vehicles $\tau(t_1)$ is *defined* as a time it takes for a vehicle to reach a road location at which the bumper of the preceding vehicle is at the time instant t_1. In single vehicle data measured at a road detector (Fig. 2.1), t_1 is the time at which the preceding vehicle leaves the detector whose location is therefore related to the location of the bumper of the preceding vehicle in the time headway definition; the time headway is equal to $\tau(t_1) = t_2 - t_1$, where t_2 is the time at which the vehicle front has been recorded at the detector. The time headway τ in (2.6) is related to the time instant t_1.

Single vehicle speeds also enable us to calculate the average (arithmetic) vehicle speed v of N vehicles passing the detector in time interval T_{av},

$$v = \frac{1}{N}\sum_{i=1}^{N} v_i, \tag{2.7}$$

where index $i = 1, 2, \ldots, N$.

The vehicle density (the number of vehicles per unit length of a road, e.g., vehicles per km) can be roughly estimated from the relation

$$\rho = \frac{q}{v}, \tag{2.8}$$

where v is the average speed. However, it should be noted that the vehicle density ρ is related to vehicles on a road section of a given length whereas the vehicle speed is measured at the location of the detector only and is averaged over the averaging time interval T_{av}. As a result, at low average vehicle speeds, the vehicle density estimated via (2.8) can lead and does usually lead to a considerable discrepancy in comparison with the real vehicle density. For a more detailed consideration of the criticism of measured data analyses associated with a considerably error in the density estimation with formula (2.8) see [5] and a recent review [6].

A road detector can also measure a macroscopic traffic variable called *occupancy*, which is defined through the formula (e.g., [1]):

$$o = \frac{T_{\text{veh}}}{T_{\text{av}}} 100\%, \tag{2.9}$$

where T_{veh} is the sum of the time intervals when the detector has measured vehicles during the time interval T_{av}:

$$T_{\text{veh}} = \sum_{i=1}^{N} \Delta t_i, \tag{2.10}$$

Δt_i is defined via (2.1).

2.2 Free Flow (F) and Congested Traffic

2.2.1 Definition of Congested Traffic

Free traffic flow (free flow for short) is usually observed, when the vehicle density in traffic is small enough. At small enough vehicle density, interactions between vehicles in free flow are negligible. Therefore, vehicles have an opportunity to move with their desired maximum speeds (if this speed is not restricted by road conditions or traffic regulations).

When the density increases in free flow, the flow rate increases too, however, vehicle interaction cannot be neglected any more. As a result of vehicle interaction in free flow, the average vehicle speed decreases with increase in density.

To illustrate these well-known features [1–3], the flow rate and density, which is calculated with formula (2.8) from the flow rate and average speed measured at a road location, are presented in the flow–density plane (points left of a dashed line FC in Fig. 2.2 (a)). In empirical traffic data, the increase in the flow rate with the density increase in free flow has a limit. At the associated *limit (maximum) point of free flow*, the flow rate and density reach their maximum values denoted by $q_{\max}^{(\text{free, emp})}$ and $\rho_{\max}^{(\text{free, emp})}$, respectively, while the average speed has a minimum value for the free flow:

$$v_{\min}^{(\text{free, emp})} = q_{\max}^{(\text{free, emp})} / \rho_{\max}^{(\text{free, emp})}. \tag{2.11}$$

These points are well-fitted by a flow–density relationship for free flow, i.e., a certain curve with a positive slope between the flow rate and density associated with averaging of measured data shown left of the dashed line FC in Fig. 2.2 (a) to one average flow rate for each density (curve F in Fig. 2.2 (b)) [1–3, 7, 8]. This flow–density relationship is called the fundamental diagram of free flow. The empirical fundamental diagram of free flow is cut off at the limit point of free flow $(\rho_{\max}^{(\text{free, emp})}, q_{\max}^{(\text{free, emp})})$ (Fig. 2.2 (b)) [1–3, 9].

To distinguish free flow points in the flow–density plane, we use in Fig. 2.2 (a, b) the dashed line FC between the origin of the flow–density plane and the limit point of free flow; the slope of the line FC is equal to the minimum speed in free flow $v_{\min}^{(\text{free, emp})}$ (2.11). Thus empirical points of free flow as well as the associated fundamental diagram lie to the left of the dashed line FC in the flow–density plane.

In empirical observations, when density in free flow increases and becomes great enough, the phenomenon of the onset of congestion is observed in this free flow: the average speed decreases abruptly to a lower speed in congested traffic:

- Congested traffic is defined as a state of traffic in which the average speed is *lower* than the minimum average speed that is still possible in free flow (e.g., [3, 10]).

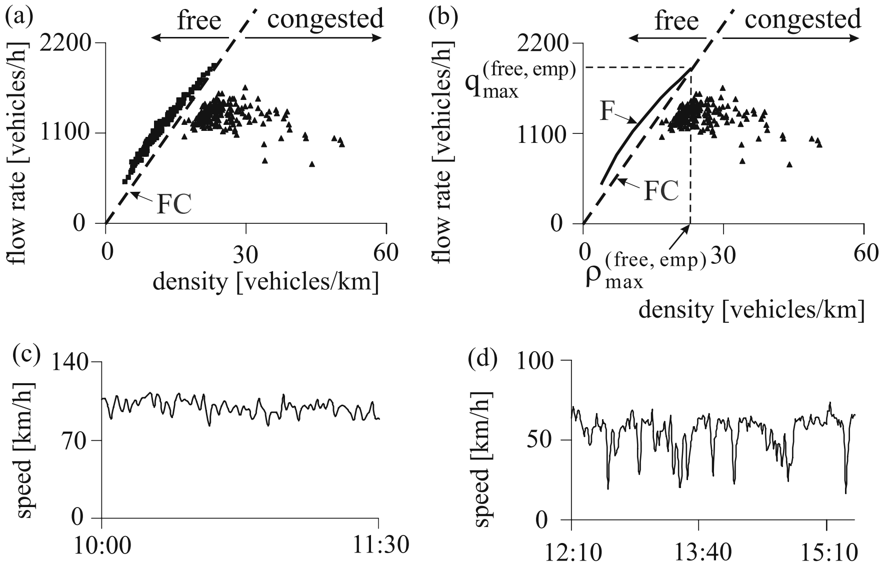

Fig. 2.2 Free flow and congested traffic (e.g., [1–3, 9, 10]). (a) Empirical data for free flow (points left of the dashed line FC) and for congested traffic (points right of the dashed line FC). (b) The fundamental diagram for free flow (curve F) and the same measured data for congested traffic as those in (a). (c, d) Vehicle speed in free flow (c) and congested traffic (d), related to points left and right of the line FC in (a), respectively. 1-min average data measured at a road location

Thus empirical points of congested traffic lie to the right of the dashed line FC in the flow–density plane[3].

Traffic congestion occurs mostly at a highway bottleneck (bottleneck for short). The bottleneck can be a result of road works, on- and off-ramps, a decrease in the number of road lanes, road curves and road gradients, bad weather conditions, accidents, etc. [1–3].

In congested traffic, a great variety of congested traffic patterns are observed [10–16]. A *congested traffic pattern* (congested pattern for short) is defined as follows.

- A congested traffic pattern is a spatiotemporal traffic pattern within which there is congested traffic. The congested pattern is separated from free flow by the downstream and upstream fronts: At the downstream front, vehicles accelerate

[3] It must be noted that the definition of congested traffic through the use of the empirical limit point on the fundamental diagram of free flow seems to be easy, however, can lead to an error in measurements of the minimum speed that is possible in free flow. This is because the exact value of this minimum speed that is possible in free flow is associated with the maximum (limit) flow rate $q_{\max}^{(\text{free, emp})}$ in free flow at which probability of traffic breakdown is equal to one (see explanations of the flow rate dependence of breakdown probability in Sect. 4.2.2). However, it is extremely difficult to find such a free flow in real measured traffic data. This comment is also related to the limit point for free flow shown in Fig. 2.2: the speed $v_{\min}^{(\text{free, emp})}$ associated with this limit point for free flow gives only an approximate value for the minimum speed that is possible in free flow.

from a lower speed within the pattern to a higher speed in free flow downstream; at the upstream front, vehicles decelerate from a free flow speed to a lower speed within the congested pattern.

In particular, one of the congested traffic patterns is a moving traffic jam [10–16]. A *moving traffic jam* (moving jam for short) is defined as follows:

- A moving jam is a localized congested traffic pattern that moves upstream in traffic flow (Fig. 2.3). Within the moving jam the average vehicle speed is very low (sometimes as low as zero), and the density is very high. The moving jam is spatially restricted by the downstream jam front and upstream jam front. Within the downstream jam front vehicles accelerate from low speed states within the jam to higher speeds in traffic flow downstream of the moving jam. Within the upstream jam front vehicles must slow down to the speed within the jam. Both jam fronts move upstream. Within the jam fronts the vehicle speed, flow rate, and density vary abruptly.

Moving jams have been studied empirically by many authors, in particular, in classic empirical works by Edie *et al.* [11–14], Treiterer *et al.* [15, 16] (Fig. 2.3), and Koshi *et al.* [10].

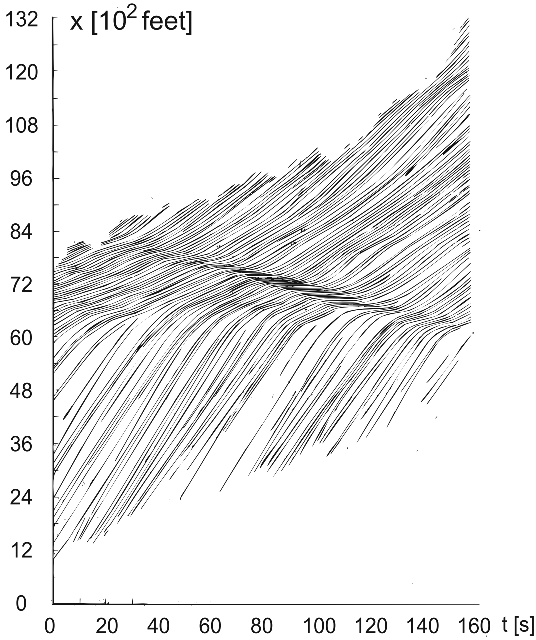

Fig. 2.3 A moving jam: dynamics of vehicle trajectories derived from aerial photography (1 feet is equal to 0.3048 m). Each of the curves in this figure shows a vehicle trajectory in the time–space plane. Taken from Treiterer [16]

2.2.2 *Traffic Breakdown*

The onset of congestion in an initial free flow is accompanied by a abrupt decrease in average vehicle speed in the free flow to a considerably lower speed in congested traffic (Figs. 2.4 and 2.5). This speed breakdown occurs mostly at highway bottlenecks and is called the breakdown phenomenon or traffic breakdown (see [9, 17–21] and earlier works referred to in the book [1] and in Chap. 2 written by Hall in [3]).

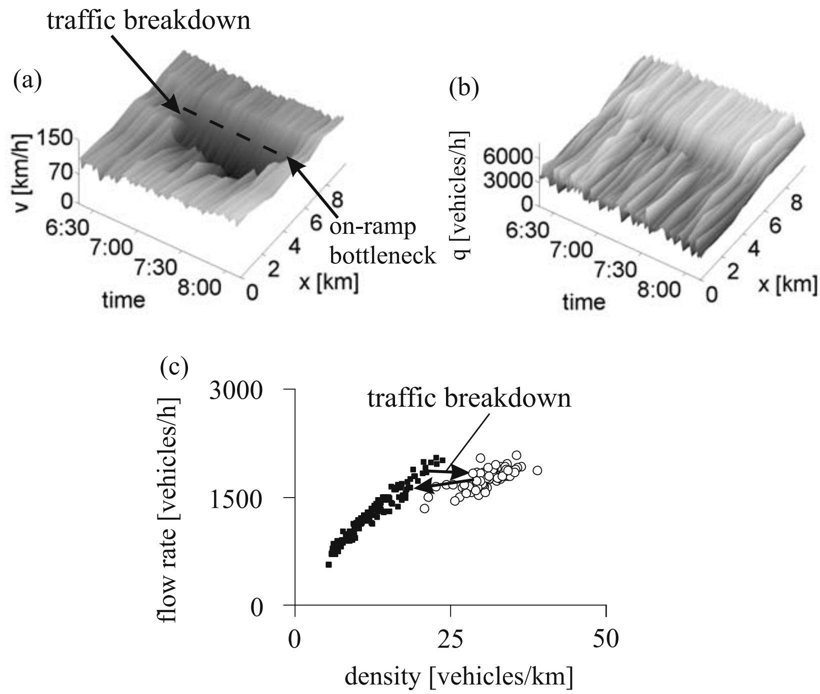

Fig. 2.4 Empirical example of traffic breakdown and hysteresis effect at on-ramp bottleneck: (a, b) Average speed (a) and flow rate (b) on the main road in space and time (note that the flow rate increase downstream of the bottleneck seen in (b) is associated with the on-ramp inflow). (c) Hysteresis effect in the flow–density plane labeled by two arrows representing traffic breakdown and return transition from congested traffic to free flow. 1-min average data. This example of traffic breakdown is qualitatively the same as many other examples observed in various countries (e.g., [9, 17–21])

The flow rate in free flow downstream of a bottleneck measured just before traffic breakdown occurs is called the *pre-discharge flow rate*. The flow rate in free flow downstream of a bottleneck after traffic breakdown has occurred at this bottleneck, i.e., the flow rate in the congested pattern outflow is called the *discharge flow rate* [17].

Hall and Agyemang-Duah have found [17] that

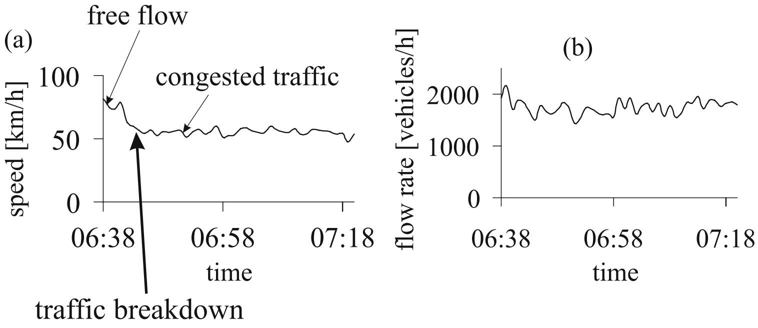

Fig. 2.5 Traffic breakdown at on-ramp bottleneck. Vehicle speed (a) and flow rate downstream of the bottleneck (b) as functions of time related to Fig. 2.4 (e.g., [9, 17–19])

- the discharge flow rate can be as great as the pre-discharge flow rate: in some cases, the discharge flow rate is smaller, however, in other cases it is greater than the pre-discharge flow rate.

2.2.3 Probabilistic Features of Traffic Breakdown

In 1995, Elefteriadou *et al.* found that traffic breakdown at a bottleneck has a probabilistic nature [18]. This means the following: at a given flow rate in free flow downstream of the bottleneck traffic breakdown can occur but it should not necessarily occur. Thus on one day traffic breakdown occurs, however, on another day at the same flow rates traffic breakdown is not observed.

Persaud *et al.* found [19] that empirical probability of traffic breakdown at a bottleneck is an increasing flow rate function (Fig. 2.6). Later such an empirical probability of traffic breakdown was also found on different highways in various countries [22–27].

Another empirical probabilistic characteristic of traffic breakdown is as follows. At given traffic parameters (weather, etc.), the flow rate downstream of an on-ramp bottleneck associated with the empirical maximum flow rate in free flow $q_{\max}^{(\text{free, emp})}$, which was measured on a specific day before congestion occurred, can be greater than the pre-discharge flow rate denoted by $q_{\text{FS}}^{(\text{B})}$ in Fig. 2.7.

After traffic breakdown has occurred, the emergent congested pattern shown in Fig. 2.4 (a, b) exists for about one hour at the bottleneck: at 7:40 free flow occurs at the bottleneck. This restoration of free flow is related to a reverse transition from congested traffic to the free flow at the bottleneck. Traffic breakdown and the reverse transition are accompanied by a well-known *hysteresis effect* and hysteresis loop in the flow–density plane: a congested pattern emerges usually at a greater flow rate downstream of the bottleneck than this flow rate is at which the congested pattern dissolves (see references in [9, 17, 20, 21]) (Fig. 2.4 (c)).