Failure of classical traffic flow theories: a critical review

B. S. Kerner

We explain that the fundamental empirical basis for automatic driving, reliable control and optimization of traffic and transportation networks is the set of empirical features of traffic breakdown at a road bottleneck. We show why generally accepted traffic and transportation theories and models are not consistent with this empirical fundament of traffic science. In particular, these classical traffic theories are as follows:

(i) the Lighthill-Whitham-Richards (LWR) theory and traffic flow models in the framework of the LWR theory (for example, Daganzo's cell transmission model) that explain traffic breakdown through a fundamental diagram of traffic flow,
(ii) General Motors (GM) class of traffic-flow models that explain traffic breakdown through traffic flow instability due to a driver reaction time (for example, the following well-known models belong to the GM model class: Gipps's model, Payne's model, Newell's optimal velocity (OV) model, Wiedemann's model (VISSIM traffic simulation tool), Bando et al. OV model, Treiber's Intelligent Driver Model, Krauß model (SUMO tool), the Aw-Rascle model),
(iii) the classical understanding of stochastic highway capacity, and
(iv) Wardrop's principles for dynamic control, assignment, and optimization of traffic and transportation networks.

In turn, this can explain why dynamics network optimization and control approaches based on these classical traffic flow theories failed by their applications in the real world. We discuss why rather that the assumption about the existence of stochastic highway capacity, at any time instant there should be the infinite number of highway capacities within a range of the flow rate between a minimum capacity and a maximum capacity as assumed in three-phase theory introduced by the author. Because the assumption about the infinite number of highway capacities is consistent with the set of the fundamental empirical features of traffic breakdown at highway bottlenecks, this can be considered a theoretical fundament for the development of reliable automatic driving, control and optimization of vehicular traffic and transportation networks. We discuss briefly some features of the three-phase theory explaining the empirical fundament of transportation science.

Keywords: empirical fundament of transportation science; traffic control and optimization; automatic driving; failure of classical traffic theories; three-phase traffic theory

Das Versagen klassischer Verkehrsfluss-Theorien: Eine kritische Betrachtung.

Wir zeigen, dass das Set der empirischen Eigenschaften eines Verkehrszusammenbruchs an einer Engstelle der Straße die fundamentale empirische Basis für automatisiertes Fahren, für zuverlässige Kontrolle und Optimierung von Verkehrs- und Transportnetzen darstellt. Wir zeigen, warum allgemein akzeptierte Theorien und Modelle des Straßenverkehrs und des Transports nicht mit diesem empirischen Fundament der Verkehrswissenschaften übereinstimmen. Im Einzelnen sind dies die folgenden klassischen Verkehrstheorien:

(i) *die Lighthill-Witham-Richards(LWR)-Theorie und Verkehrsfluss-Modelle im Rahmen der LWR-Theorie (zum Beispiel Daganzos Cell Transmission-Modell), die Verkehrszusammenbrüche durch ein Fundamentaldiagramm des Verkehrsflusses erklären,*
(ii) *die Klasse der General-Motors(GM)-Verkehrsflussmodelle, die Verkehrszusammenbrüche durch eine Instabilität des Verkehrsflusses aufgrund der Reaktionszeit des Fahrers erklären (zum Beispiel gehören die folgenden bekannten Modelle zur Klasse der GM-Modelle: Gipps-Modell, Payne-Modell, Newells Optimal Velocity(OV)-Modell, Wiedemann-Modell (Verkehrssimulations-Programm VISSIM), das OV-Modell von Bando et al., Treibers Intelligent Driver-Modell, Krauß-Modell (Verkehrssimulationsprogramm SUMO), das Aw-Rascle-Modell),*
(iii) *das klassische Verständnis der stochastischen Kapazität einer Straße,*
(iv) *das Wardrop-Prinzip für die dynamische Kontrolle, dynamische Verkehrsumlegung und Optimierung von Verkehrs- und Transportnetzwerken.*

Aus dem Scheitern der klassischen Verkehrstheorien folgt, warum auch die Ansätze zur Optimierung und Kontrolle von dynamischen Netzwerken, die auf diesen klassischen Verkehrsflusstheorien basieren, in den Anwendungen der realen Welt scheitern. Wir diskutieren, warum es anstelle der Annahme der Existenz eines bestimmten Wertes der stochastischen Kapazität einer Straße zu jedem Zeitpunkt stattdessen zu jedem Zeitpunkt eine unendliche Anzahl von Kapazitäten gibt, die in einem Bereich des Flusses zwischen einer minimalen und einer maximalen Kapazität liegen, so wie in der Drei-Phasen-Theorie angenommen wird. Weil die Annahme der Existenz einer unendlichen Anzahl von Kapazitäten einer Straße zu jedem Zeitpunkt in Übereinstimmung mit dem Set der fundamentalen empirischen Eigenschaften eines Verkehrszusammenbruchs an einer Engstelle der Straße steht, kann dies als theoretisches Fundament angesehen werden für die Entwicklung von zuverlässigem automatisierten Fahren, von Kontrolle und Optimierung des Fahrzeugverkehrs und von Transportnetzwerken. Wir diskutieren kurz einige Eigenschaften der Drei-Phasen-Verkehrstheorie, die das empirische Fundament der Verkehrswissenschaften erklären.

Schlüsselwörter: empirisches Fundament der Verkehrswissenschaften; Verkehrskontrolle und -optimierung; automatisiertes Fahren; Versagen der klassischen Verkehrstheorien; Drei-Phasen-Verkehrstheorie

Received June 29, 2015, accepted September 12, 2015, published online October 1, 2015

Kerner, Boris S., Physics of Transport and Traffic, University Duisburg-Essen, 47048 Duisburg, Germany (E-mail: boris.kerner@uni-due.de)

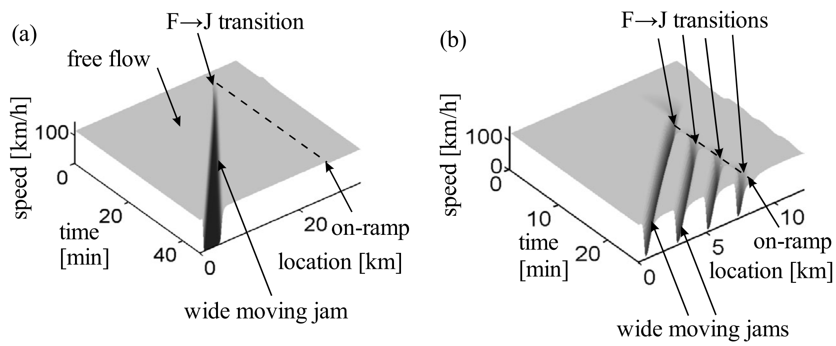

Fig. 1. Simulations of traffic breakdown at on-ramp bottleneck in the GM model class: Speed in space and time during the spontaneous emergence of wide moving jam (J) in initial free flow (F) at on-ramp bottleneck (F → J transition). The flow rate on the main road upstream of the bottleneck q_{in} is larger in (a) than that in (b), whereas on-ramp inflow q_{on} is smaller in (a) than that in (b) at the same other model parameters (see Sect. 10.3 of the book [32] for more detail)

1. Introduction

As explained in a recent critical review [1], generally accepted fundamentals and methodologies of traffic and transportation theory have failed by their applications for traffic network optimization and control in the real world. In comparison with the above-mentioned review with about 540 references, this brief critical review is mostly devoted to formulation and critical discussion of empirical and theoretical fundamentals, which can be used for the development of reliable automatic driving, control and optimization of vehicular traffic and transportation networks.

Therefore, objectives of this critical review are as follow:

(i) We show why the fundamental empirical basis for automatic driving, reliable control and optimization of traffic and transportation networks is the set of empirical features of traffic breakdown at a road bottleneck.
(ii) We explain why the classical generally accepted traffic flow models and associated well-known traffic simulation tools cannot be used for analyses of automatic driving, reliable control and optimization of vehicular traffic and transportation networks.
(iii) We present results of three-phase traffic theory that can show these empirical features of traffic breakdown. Therefore, three-phase traffic flow models can be used as reliable tools for the development of automatic driving vehicles that improve traffic safety and decrease the probability of traffic congestion in traffic and transportation networks.

2. Generally accepted fundamentals and methodologies of traffic and transportation theory

Traffic researchers have developed a huge number of traffic theories for optimization and control of traffic and transportation networks. In particular, generally accepted fundamentals and methodologies of traffic and transportation theory are as follow:

i. The Lighthill-Whitham-Richards (LWR) model introduced in 1955–1956 [2, 3]. Daganzo introduced a cell-transmission model (CTM) that is consistent with the LWR model [4, 5]. Currently, Daganzo's CTM is widely used for simulations of traffic and transportation networks (see references in [1]).
ii. A traffic flow instability that causes a growing wave of a local reduction of the vehicle speed. This classical traffic flow instability was introduced in 1959–1961 in the General Motors (GM) car-following model by Herman, Gazis, Montroll, Potts, Rothery, and Chandler [6–8] (see also other references in the book [9]).

With the use of very different mathematical approaches, this classical traffic flow instability [6–8] has been incorporated in a huge number of traffic flow models that can be considered belonging to the GM model class. This is because (as found firstly by Kerner and Konhäuser in [10, 11]) in all these very different traffic flow models the traffic flow instability that causes a growing wave of a local reduction of the vehicle speed leads to a moving jam (J) formation in free flow (F) (called F → J transition; see Fig. 1).

Examples of the well-known models that belong to the GM model class are as follows: Gipps's model, Payne's model, Newell's optimal velocity (OV) model, Wiedemann's model (used in VISSIM traffic simulation tool), Whitham's model, the Nagel-Schreckenberg (NaSch) cellular automaton (CA) model, Bando et al. OV model, Treiber's IDM, Krauß model (used in SUMO tool), the Aw-Rascle model and many other well-known microscopic and macroscopic traffic-flow models. These models are the basis of a number of traffic simulation tools widely used by traffic engineers and researchers (see e.g., references in reviews [9, 12–20]).

iii. The understanding of highway capacity as a *particular* value. This understanding of road capacity was probably introduced in 1920–1935 (see the classical paper by Greenshields [21] and references in [12, 15, 16, 20]). Recently, due to empirical results of [22–29] it has been assumed this the particular highway capacity is a stochastic value (see Appendix A).
iv. Wardrop's user equilibrium (UE) and system optimum (SO) principles for traffic and transportation network optimization and control introduced in 1952 [30]. The Wardrop's UE and SO principles are the basis for a huge number of models for dynamic traffic assignment, control and optimization of traffic and transportation networks (see references in [1, 31]).

3. The fundamental empirical basis for automatic driving as well as for reliable control and optimization of vehicular traffic and transportation networks

Vehicular traffic is a spatiotemporal process because it occurs in space and time. Traffic and transportation networks are usually very complex. Therefore, it is not surprising that in empirical studies of traffic data measured in vehicular traffic a diverse variety of empirical spatiotemporal traffic phenomena have been discovered. Obviously, each of the traffic and transportation theories and models can explain some real traffic phenomena and each of the models exhibits a limited region of the applicability for the explanation of real traffic

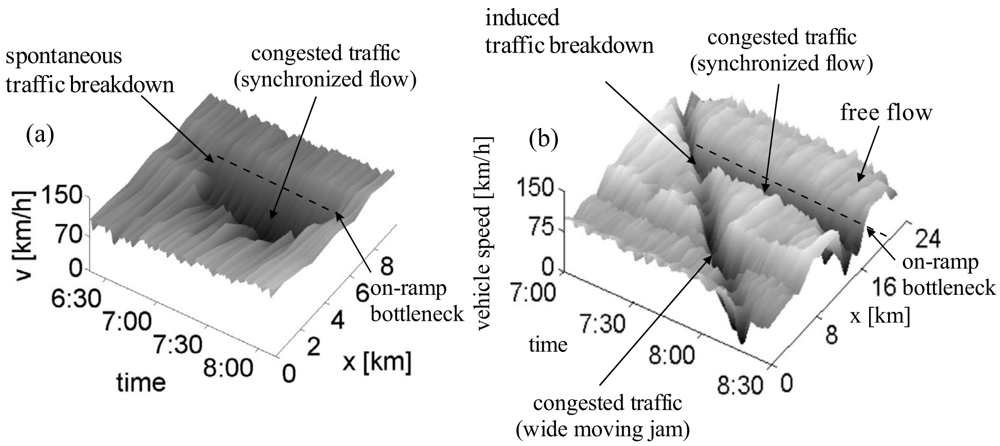

Fig. 2. Examples for empirical spontaneous (a) and empirical induced traffic breakdown (b) at highway bottlenecks (1 min averaged field data measured by road detectors on freeway A5-South in Germany). Taken from [33]. This empirical data explains the qualitative difference between two phases in congested traffic—wide moving jam and synchronized flow as follows. The downstream front of the wide moving jam (J) in (b) propagates upstream of the bottleneck with a constant mean velocity. In contrast, the downstream front of synchronized flow is fixed at the bottleneck. This different behavior of the downstream front of congested traffic is the basis for the definition of the wide moving jam and synchronized flow phases in congested traffic (more detailed explanations as well as the motivation of these definitions see in Sects. 2.3–2.7 of the book [32]; the concept of synchronized flow has also been explained in Sect. 4.2 of the book [33])

and transportation phenomena. Therefore, the following question arises:

- Whether is there an empirical traffic phenomenon that can be considered the fundamental empirical basis of transportation science?

Users of traffic and transportation networks would expect that through the use of traffic control, dynamic traffic assignment and other methods of dynamic optimization traffic breakdown can be prevented, i.e., free flow can be maintained in the network. This is because due to traffic breakdown congested traffic occurs in which travel time, fuel consumption as well as other travel costs increased considerably in comparison with travel costs in free flow.

Therefore, any traffic and transportation theory, which is claimed to be a basis for the development of reliable methods and strategies for dynamic traffic assignment as well as network optimization and control should be consistent with the fundamental empirical features of traffic breakdown at a road bottleneck.

The fundamental empirical basis for automatic driving as well as for reliable control and optimization of vehicular traffic and transportation networks is the set of empirical features of traffic breakdown at a road bottleneck. Consequently, we can also make the following conclusion:

- Traffic and transportation theories, which are not consistent the set of the fundamental empirical features of traffic breakdown at a bottleneck, cannot be applied for the development of reliable management, control, and organization of traffic and transportation networks.

4. The set of fundamental empirical features of traffic breakdown at highway bottlenecks

The set of fundamental empirical features of traffic breakdown at a highway bottleneck, which is found from a study of traffic breakdown at road bottlenecks during many different days (and years) of traffic breakdown observations, is as follows [32, 33]:

1. Traffic breakdown at a highway bottleneck is a local phase transition from free flow (F) to congested traffic whose downstream front is usually fixed at the bottleneck location (see, e.g., [12, 15, 20, 22–29] and references there). In three-phase traffic theory, such congested traffic is called synchronized flow (S) [32, 33] (Fig. 2).
2. At the same bottleneck, traffic breakdown can be either spontaneous (Fig. 2(a)) or induced (Fig. 2(b)) [32, 33] (see Appendix B).
3. As found firstly by Elefteriadou et al. in 1995 [22], traffic breakdown exhibits a probabilistic nature: At the same bottleneck, traffic breakdown is observed on different days at very different flow rates. In 1998, Persaud et al. [25] found that probability of traffic breakdown is an increasing flow rate function. This result has been confirmed in empirical studies of data measured in different countries [25–29, 34].
4. There is a well-known hysteresis phenomenon associated with traffic breakdown and a return transition to free flow (e.g., [12, 20, 23–29]).

5. Explanation of failure of classical traffic and transportation theories

As emphasized in [1], there are many achievements of the generally accepted fundamentals and methodologies of traffic and transportation theory (Sect. 2), which have made a great impact on the understanding of many traffic phenomena. Because of these achievements of generally accepted classical traffic and transportation theories, a question arises:

- Why does the author state in [1] that the generally accepted classical traffic and transportation theories are not consistent with the set of empirical features of traffic breakdown and, therefore, they are not applicable for a reliable description of traffic breakdown, capacity, effect of automatic driving on real traffic flow, traffic control, and optimization of real traffic and transportation networks?

The failure of the generally accepted classical traffic flow theories is explained as follows [1]:

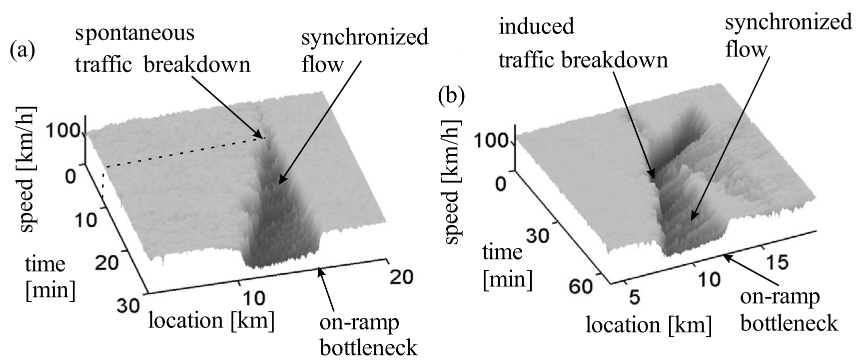

Fig. 3. Simulations of spontaneous (a) and induced (b) breakdown at on-ramp bottleneck with a microscopic stochastic traffic flow model in the framework of three-phase traffic theory [38]

1. The LWR-theory [2–5] failed because this theory cannot show induced traffic breakdown observed in real traffic (Fig. 2(b)) (see the proof of this statement in Sect. 4 of [1]).
2. Traffic flow models of the GM model class (see references in [9, 12–14, 17–20]) failed because traffic breakdown in the models of the GM class is an F → J transition (Fig. 1): Due to traffic breakdown, a wide moving jam(s) appears spontaneously in an initially free flow at a bottleneck (Fig. 1). In contrast with this model result, real traffic breakdown is a phase transition from free flow (F) to synchronized flow (S) (F → S transition): Rather than a wide moving jam(s), due to traffic breakdown in real traffic, synchronized flow occurs whose downstream front is fixed at the bottleneck (Fig. 2) (a more detailed explanation of the definitions and features of the F → S and F → J transitions can be found, respectively, in Chap. 3 and Sect. 10.3 of the book [32]).
3. The understanding of highway capacity as a particular value [12, 15, 16, 21–29] failed because this assumption about the nature of highway capacity contradicts the empirical evidence that traffic breakdown can be induced at a highway bottleneck (Fig. 2(b)) (see Appendixes A–D).
4. Dynamic traffic assignment or/and any kind of traffic optimization and control based on Wardrop's SO or UE principles (see references in [1, 31]) failed because of possible random transitions between the free flow and synchronized flow at highway bottlenecks. Due to such random transitions, the minimization of travel cost in a traffic network is not possible.

This can explain why network optimization and control approaches based on these fundamentals and methodologies failed by their applications in the real world. Even several decades of a very intensive effort to improve and validate network optimization models have no success. Indeed, there can be found no examples where on-line implementations of the network optimization models based on these fundamentals and methodologies could reduce congestion in real traffic and transportation networks.

This is due to the fact that the fundamental empirical features of traffic breakdown at highway bottlenecks have been understood only during last 20 years. In contrast, the generally accepted fundamentals and methodologies of traffic and transportation theory have been introduced in the 50–60 s.

Thus the scientists whose ideas led to these classical fundamentals and methodologies of traffic and transportation theory (see Sect. 2 above) could not know the set of empirical features of real traffic breakdown. It should be noted that many of the diverse driver behavioral characteristics related to real traffic as well as some of the mathematical approaches to traffic flow modeling, which have been discovered in classical approaches to traffic flow theory, are also used in three-phase traffic theory and associated microscopic traffic flow models (for more details, see Sect. 11 of [1]).

6. Basic theoretical fundament for the development of reliable control and optimization of traffic and transportation networks

To explain the set of the fundamental empirical features of traffic breakdown at network bottlenecks, the author has introduced three-phase traffic flow theory [32, 33, 35–37].

- The main reason for the three-phase traffic theory is the explanation of the set of the fundamental empirical features of traffic breakdown at highway bottlenecks.

In three-phase traffic theory, an F → S transition explains traffic breakdown at a highway bottleneck: The terms "F → S transition" and "traffic breakdown" are synonyms. The F → S transition (traffic breakdown) occurs in metastable free flow (Figs. 3 and 4) [33] (see explanation of the term "metastable free flow with respect to an F → S transition" in Appendixes B–D). The metastability of free flow explains both spontaneous (Fig. 3(a)) and induced (Fig. 3(b)) traffic breakdowns leading to the emergence of synchronized flow at the bottleneck (empirical features 1 and 2 of traffic breakdown of Sect. 4).

The theoretical probability of spontaneous traffic breakdown at the bottleneck found firstly from simulations of a microscopic stochastic three-phase traffic flow model [39] (Fig. 4(a)) is a growing flow-rate function as discovered by Persaud et al. [25] in field data measured by road detectors (empirical feature 3 of traffic breakdown of Sect. 4). The theoretical dependence of the probability $P^{(B)}(q_{\rm sum})$ of spontaneous traffic breakdown at the on-ramp bottleneck on the flow rate $q_{\rm sum}$ in free flow at the bottleneck is well fitted by a function [39]

$$P^{(B)}(q_{\rm sum}) = \frac{1}{1+\exp[\alpha\,(q_P - q_{\rm sum})]}, \qquad (1)$$

where $q_{\rm sum} = q_{\rm in} + q_{\rm on}$ is the flow rate downstream of the bottleneck, parameters α and q_P depend on the on-ramp inflow rate $q_{\rm on}$ and a time interval within which traffic breakdown is studied, $q_{\rm in}$ is the flow rate in free flow on the main road upstream of the bottleneck. In other words, in accordance with empirical data, in simulations has been found that the on-ramp inflow rate $q_{\rm on}$ and the flow rate in free flow on the main road upstream of the bottleneck $q_{\rm in}$ have different contributions to traffic breakdown (see Fig. 18(b, c) of [39]). Qualitatively the same growing flow-rate function for the

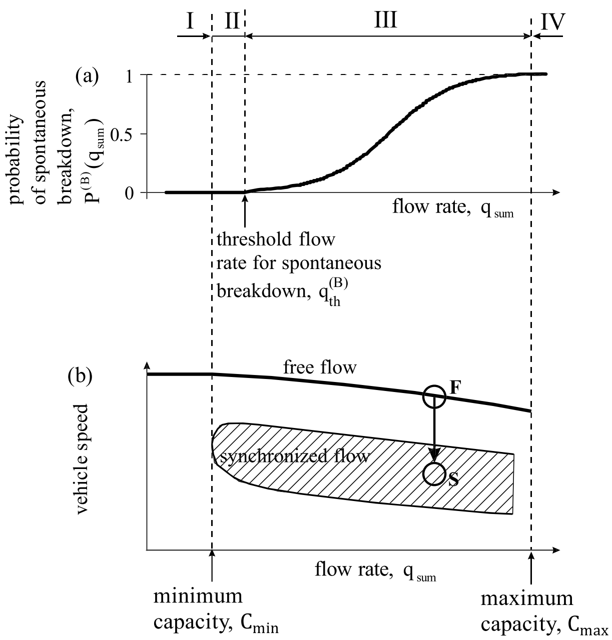

Fig. 4. Explanations of the fundamental empirical features of traffic breakdown at on-ramp bottleneck based on the assumption of three-phase traffic theory about the metastability of free flow at the bottleneck with respect to an F → S transition (taken from [1]): (a) Simulations of the probability of spontaneous traffic breakdown at the bottleneck on a single-lane road taken from [39]. (b) Qualitative Z-speed-flow-rate characteristic for traffic breakdown; F—free flow, S—synchronized flow. Figure (b) has been qualitatively drawn in accordance with simulations of the Z-speed-flow-rate characteristic for traffic breakdown at on-ramp bottleneck shown in Fig. 3.17(b) of the book [32] in which the flow rate in free flow downstream of the bottleneck $q_{sum} = q_{in} + q_{on}$ changes through a change in the on-ramp inflow q_{on} at a given flow rate in free flow upstream of the bottleneck q_{in}; in these simulations, due to the increase in q_{on} at constant q_{in}, average synchronized flow speed decreases slightly (in Fig. 3.17(b) of the book [32], the synchronized flow speed is related to a virtual road detector located 200 m upstream of the on-ramp merging region)

breakdown probability (1) has later been found in field data measured by road detectors in [27–29].

The possibility of empirical induced traffic breakdown at a bottleneck (Fig. 2(b)) leads to the following conclusion of three-phase traffic theory: At the same flow rate on a network link, traffic flow at the bottleneck can be either in the free flow phase (F) or in the synchronized flow phase (S) (Fig. 4(b)). In three-phase theory, this empirical fact is responsible for the existence at any time instant of the range of the infinite number of highway capacities, which are within the flow-rate range between a minimum highway capacity C_{min} and a maximum highway capacity C_{max} (Fig. 4); within this flow-rate range, traffic breakdown can be induced at the bottleneck. Thus, the theoretical fundament resulting from three-phase traffic theory for the development of reliable control and optimization of traffic and transportation networks is as follows.

- *At any time instant*, there are the infinite number of the flow rates in free flow at a bottleneck at which traffic breakdown can be induced at the bottleneck. These flow rates are the infinite number of the capacities of free flow at the bottleneck. The range of these capacities of free flow at the bottleneck is limited by the minimum highway capacity C_{min} and the maximum highway capacity C_{max} (Fig. 4).

The sense of the infinite number of highway capacities of free flow at the bottleneck is as follows (see also Appendix C). When the flow rate in free flow at the bottleneck is within the flow rate range

$$C_{min} \le q_{sum} < C_{max}, \tag{2}$$

free flow is in a *metastable* state with respect to traffic breakdown (F → S transition) at the bottleneck. This means that traffic breakdown can occur. Therefore, all flow rates satisfying conditions (2) are highway capacities.

In general, there can be four ranges I, II, III, and IV of the flow rate (Fig. 4) within which free flow at a road bottleneck exhibits qualitatively different features with respect to traffic breakdown [32, 33].

In the range I of the flow rate related to condition

$$q_{sum} < C_{min}, \tag{3}$$

free flow is stable with respect to traffic breakdown (Fig. 4). This means that highway capacity does not reach.

In the range II of the flow rate related to condition

$$C_{min} \le q_{sum} < q^{(B)}_{th}, \tag{4}$$

free flow is metastable with respect to traffic breakdown; however, the breakdown can be induced only. This is because under condition (4) the probability of spontaneous breakdown during a given time interval is equal to zero: $P^{(B)} = 0$. In conditions (4), $q_{sum} = q^{(B)}_{th}$ (Fig. 4(a)) is a threshold flow rate for spontaneous traffic breakdown at which the breakdown probability $P^{(B)}(q^{(B)}_{th})$ is very small but it is still larger than zero.

In the range III of the flow rate related to condition

$$q^{(B)}_{th} \le q_{sum} < C_{max}, \tag{5}$$

free flow is also metastable with respect to traffic breakdown. However, in contrast with range II (conditions (4)) traffic breakdown can occur spontaneously at the bottleneck during a given time interval with the probability $0 < P^{(B)}(q_{sum}) < 1$. Naturally, under conditions (5) the breakdown can also be induced.

In the range IV of the flow rate related to condition

$$q_{sum} \ge C_{max}, \tag{6}$$

free flow can be considered *unstable* with respect to traffic breakdown. This is because the breakdown does occur with probability $P^{(B)} = 1$ during a time interval that is the shorter, the more the flow rate exceeds the maximum capacity.

Recently, the theoretical conclusion about the existence of the infinite number of road capacities as well as conditions (2)–(6) have been generalized for a city bottleneck due to traffic signal [40–42]. In particular, it has been found that the flow rate function of the probability of traffic breakdown (transition from under-saturated to oversaturated traffic) at the signal is also related to that shown in Fig. 4(a) [40–42]:

$$P^{(B)}(\bar{q}_{in}) = \frac{1}{1 + \exp[\beta\,(q_s - \bar{q}_{in})]}, \tag{7}$$

where $\bar{q}_{in}$ is the average arrival flow rate at the approach, parameters β and q_s depend on signal characteristics.

However, in accordance with three-phase traffic theory of city traffic introduced by the author [40–42], synchronized flow patterns occurring in undersaturated traffic flow at traffic signal are quite different from those at highway bottlenecks. As a result, rather than Fig. 4(b) that is valid for highway bottlenecks *only*, qualitatively different characteristics for a study of the metastability of undersaturated city traffic at the signal should be used.

It must be emphasized that for an empirical study of city traffic phenomena in undersaturated city traffic (metastability, synchronized flow, infinity number of signal capacities) theoretically predicted in [40–42], spatiotemporal distributions of the flow rate and speed should be measured both upstream and downstream of the signal over many days in which traffic breakdown (transition from under- to oversaturated traffic) are observed.

Unfortunately, to the knowledge of the author, such field data measured in city traffic is not available. In other words, the set of empirical features of the breakdown at the signal could not still be studied. Therefore, in contrast with highway traffic (Sect. 2 and Appendix B), there is *no* empirical proof of the theoretical result about the infinite number of capacities of traffic signal [41, 42]. For this reason, a more detailed consideration of probabilistic theory of city traffic [42] as well as of a comparison of this stochastic theory with classical two-regime models for under- and oversaturated traffic made in [43] are out of scope of this brief review.

It should be noted that a consideration of features of synchronized flow resulting from the breakdown both in highway traffic [32, 33] and in oversaturated city traffic [44, 45] is also out of scope of this review. This is because this review is solely devoted to a brief consideration of the set of fundamental empirical features of traffic breakdown occurring in free flow at highway bottlenecks as well as to the impact of these empirical features of the breakdown on theoretical fundamentals for reliable control and optimization of vehicular traffic and transportation networks.

7. Incommensurability of three-phase traffic theory and classical traffic-flow theories

Due to the criticism of classical traffic-flow theories made in Sect. 4, a question arises:

- May some of the classical traffic-flow theories be relatively easily adjusted to take into account the empirical evidence of the induced transition from free flow to synchronized flow and the flow-rate dependence of the breakdown probability?

The explanation of traffic breakdown at a highway bottleneck by an F → S transition in a metastable free flow at the bottleneck is the basic assumption of three-phase traffic theory (Fig. 4) [32, 33, 35–37]. *None* of classical traffic-flow theories (see for review, e.g., [9, 12–20]) incorporates an F → S transition in a metastable free flow at the bottleneck.

For this reason, the classical traffic-flow models cannot describe the F → S phase transition in metastable free flow at highway bottleneck. However, the transition does explain the empirical evidence of the induced transition from free flow to synchronized flow and the flow-rate dependence of the breakdown probability.

In accordance with the classical book by Kuhn [46], this shows *the incommensurability* of three-phase traffic theory and the classical traffic-flow theories (for more detail, see [47]):

- The existence in three-phase traffic theory of the minimum highway capacity C_{min} at which traffic breakdown (F → S phase transition) can still be induced at a highway bottleneck has *no* sense for classical traffic flow theories.

The term *incommensurability* has been introduced by Kuhn in his classical book [46] to explain a paradigm shift in a scientific field.

It must also be noted that the existence of these two phases F and S (Fig. 4) does not result from the stochastic nature of traffic: Even if there were no stochastic processes in vehicular traffic, the states F and S do exist at the same flow rate. For this reason, stochastic approaches to traffic control (see, e.g., [48, 49]), which do not assume a possibility of an F → S phase transition in metastable free flow and, respectively, the existence of the flow range between the minimum and maximum capacities of three-phase traffic theory, cannot resolve the above-discussed problem of the inconsistence of classical traffic theories with the set of empirical features of real traffic breakdown.

However, it should be noted that the stochastic nature of traffic influences crucially on the probability of random transitions between the phases F and S. At a given flow rate, this probability can change in several orders of magnitude when stochastic characteristics of traffic change.

8. Future reliable control and optimization of vehicular traffic with the use of three-phase traffic theory

Thus within the flow rate range (5), traffic breakdown at a highway bottleneck can occur with some probability regardless of traffic control. This explains the criticism of generally accepted methods of traffic control of Sect. 5. Therefore, a question arises:

- How will the three-phase traffic theory assist in providing reliable control and optimization of vehicular traffic?

The three-phase traffic theory provides the following future directions for traffic control and optimization theory:

a) the minimization of breakdown probability in free flow at network bottlenecks based on the breakdown minimization principle (BM) principle for the control and optimization of transportation networks [50]. The BM principle should be applied for those parts of a traffic network that are not influenced by congestion together with
b) a spatial limitation or/and dissolution of congestion in congested parts of the network.

However, a consideration of the BM principle [50] and methods for the spatial limitation or/and dissolution of congestion at bottlenecks based on three-phase traffic theory [32, 33, 51–54] is out of scope of this brief review.

9. Control of congested traffic

Up to now we have discussed empirical traffic breakdown as the empirical fundament for automatic driving as well as for reliable control and optimization of transportation networks. However, as mentioned in Sect. 8, for reliable network optimization and control, one can apply the minimization of the probability of traffic breakdown in the network with the BM principle that should be combined with a spatial limitation of congestion growth and/or congestion dissolution in congested network links.

Moreover, in many real traffic networks there are not enough alternative routes to avoid traffic congestion at large enough traffic demand. In these cases, it can be expected that even if traffic congestion cannot be avoided in some parts of traffic networks, nevertheless, the application of ITS can change characteristics of traffic congestion with the objective to increase traffic safety and comfort while moving in congested traffic. Therefore, the following questions arise:

1. What are empirical features of congested patterns that can be influenced for reliable spatial limitation of congestion growth and/or for congestion dissolution?
2. How can driver behavior change features of congested patterns with the aim to increase safety and comfortable driving?

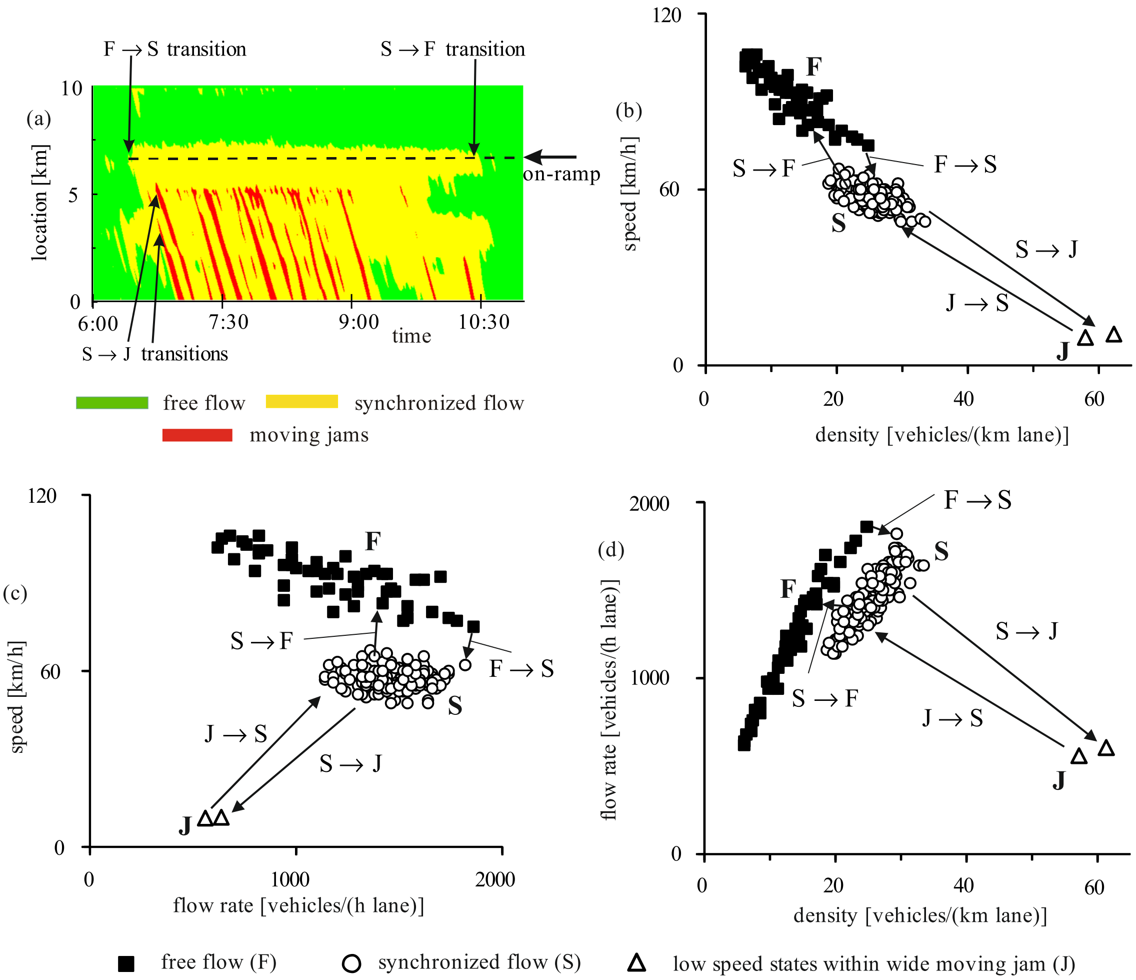

Fig. 5. Empirical sequence of F → S → J phase transitions in real traffic [55, 56]: (a) Speed in space and time within a congested pattern occurring at on-ramp bottleneck (1-min averaged data measured on March 23, 1998 on freeway A5-South in Germany with road detectors installed along a three-lane road); data is presented in space and time with averaging method described in Sect. C.2 of [57] (*arrows* F → S and S → F mark symbolically, respectively, the F → S and S → F transitions at the location of on-ramp bottleneck, *arrows* S → J mark symbolically S → J transitions related to the emergence of two first wide moving jams within synchronized flow). (b–d) Empirical double Z-characteristics, which are the presentation of phase transitions shown in (a) in the speed–density (b), speed–flow-rate (c) and flow–density planes (d) (points J in (b–d) are related to the second of the wide moving jams in (a))

3. What are vehicle systems for automatic driving as well as other ITS-applications that can help to increase safety and comfort while driving in congested traffic?

The three-phase traffic theory has answered these questions [32, 33]. In particular, in 1998 based on an analysis of real field traffic data measured on German highways the author found out that one of the most important features of spatiotemporal complexity of traffic congestion is a sequence of F → S → J phase transitions (Fig. 5) [55] that are as follows. Firstly, an F → S transition (traffic breakdown) occurs at a highway bottleneck as discussed in Sect. 6 above. Synchronized flow propagates upstream of the bottleneck. Within the emergent synchronized flow, at some distance upstream of the bottleneck moving jams emerge spontaneously (called as an S → J transition) (Fig. 5).

Both the F → S transition and the S → J transition, i.e., the sequence of F → S → J transitions (Fig. 5(a)), which occur in the reality in space and time, can alternatively be presented either in the speed–density plane (Fig. 5(b)), or in the speed–flow-rate plane (Fig. 5(c)), or else in the flow–density plane (Fig. 5(d)) by *double Z-characteristic* for phase transitions [33]. In other words, arrows F → S and S → J shown in these planes (Fig. 5(b–d)) mark, respectively, the presentation of the F → S transition and the S → J transition associated with real F → S transition and the S → J transition shown in Fig. 5(a). Obviously, there are also return S → F and J → S transitions between the three traffic phases (Fig. 5).

10. Effect of automatic driving on traffic flow

Traffic flow simulations should be able to answer the following questions, which arise due to the development of automatic driving vehicles:

1. What is the effect of automatic driving vehicles on traffic flow consisting of usual (non-automatic) vehicles?
2. What are features of vehicle systems for automatic driving that can improve traffic safety and decrease the probability of traffic congestion?

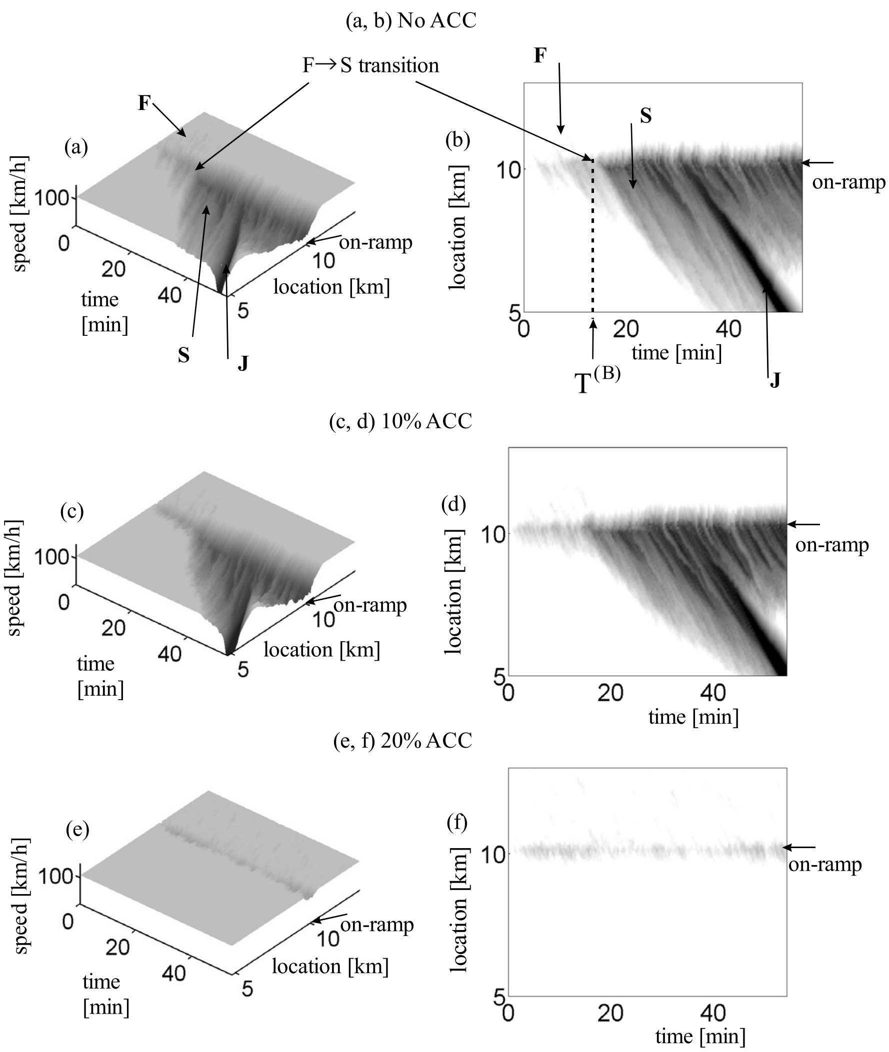

Fig. 6. Simulations of the effect of adaptive vehicle cruise control (ACC) on traffic flow on single-lane road with an on-ramp bottleneck made with Kerner-Klenov microscopic stochastic traffic flow model [64–66] in the framework of three-phase traffic theory. Vehicle speed in space and time (a, c, e) and the same speed data presented by regions with variable shades of gray (b, d, f) (in *white regions* the speed is equal to 105 km/h, in *black regions* the speed is equal to zero). (a, b) No ACC vehicles. (c, d) 10 % ACC vehicles. (e, f) 20 % ACC vehicles. Explanations of ACC used for simulations is given in Sect. 23.6 of the book [33]; desired time gap $\tau_d^{(ACC)}$ that is a given parameter of ACC is equal to 1.1 s; other parameters of ACC are the same as those in caption of Fig. 23.18 of the book [33]. *Arrows* F → S in (a, b) mark the F → S transition (traffic breakdown) at the location of on-ramp bottleneck. Flow rate in free flow upstream of the bottleneck is equal to $q_{in} = 2000$ vehicles/h, the flow rate to the on-ramp is equal to $q_{on} = 320$ vehicles/h. F—free flow, S—synchronized flow, J—wide moving jam

As explained in this review, to answer these questions, traffic flow models used for simulations should be able to explain the set of empirical features of traffic breakdown at a road bottleneck. Because the classical generally accepted traffic flow models cannot show these empirical features, the application of these models and associated simulation tools leads to incorrect conclusions that cannot be used for the development of systems for the future automatic driving vehicles.

For this reason, studies the effect of adaptive vehicle cruise control (ACC) and other vehicle systems for automatic driving on traffic flow with well-known traffic flow models and simulation tools like VISSIM (Wiedemann model), SUMO (Krauß model) as well as all other traffic simulation tools based on classical traffic flow models (see Sect. 2), which have been made and/or reviewed, for example, in [19, 58–63], *are invalid* for the real word. Therefore, simulation approaches of [19, 58–63] lead to *incorrect conclusions* about the impact of ACC vehicles as well as of automatic driving on real traffic flow.

As explained in Sect. 23.6 of the book [33] and Sect. 9.6 of the book [32], to perform a reliable simulations of the effect of ACC and other vehicle systems for automatic driving on traffic flow on traffic flow, a three-phase traffic flow model, which can explain the set of the empirical features of traffic breakdown at highway bottlenecks (Sect. 4), is needed.

One of such three-phase traffic flow models is Kerner-Klenov stochastic microscopic three-phase traffic flow model [64–66]. One of the results of simulations of the effect of ACC on traffic flow made with this model is shown in Figs. 6 and 7.

At chosen flow rates q_{in} and q_{on} (Fig. 6), in traffic flow without ACC vehicles, traffic breakdown (F → S transition) occurs at an on-ramp bottleneck after a random time delay for traffic breakdown

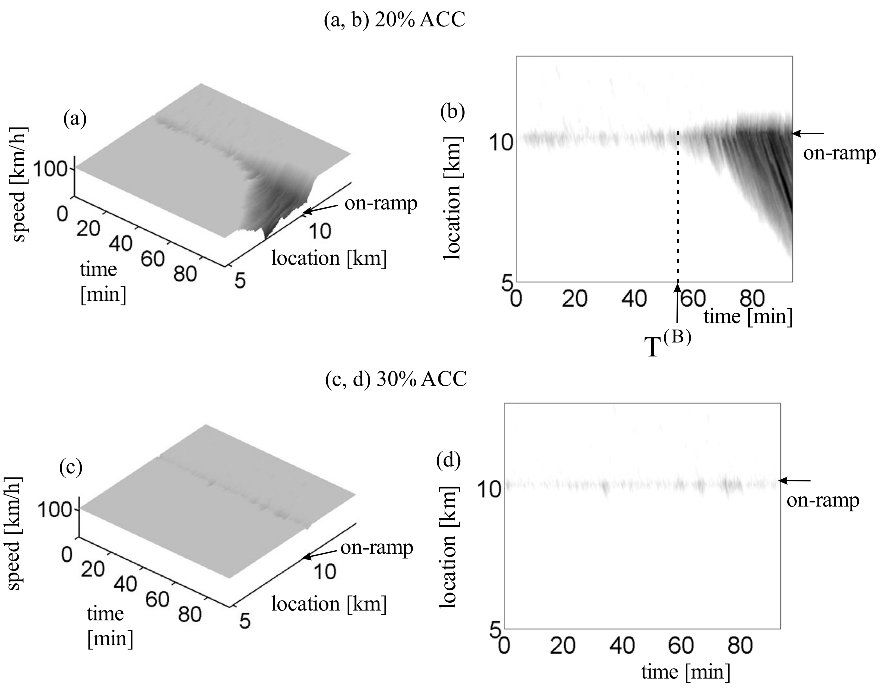

Fig. 7. Simulations of the effect of ACC on traffic flow on single-lane road with an on-ramp bottleneck in the framework with three-phase traffic theory (continue of Fig. 6). Vehicle speed in space and time (a, c) and the same speed data presented by regions with variable shades of gray (b, d) (in *white regions* the speed is equal to 105 km/h, in *black regions* the speed is equal to zero). (a, b) 20 % ACC vehicles with a longer time of observation of traffic flow than that in Fig. 6(e, f). (c, d) 30 % of ACC vehicles. Other parameters and explanations are the same as those in Fig. 6

$T^{(B)}$ (Fig. 6(a, b)) as this is well-known in three-phase traffic theory and empirical observations [32–34].

If there are 10 % ACC vehicles with quick speed adaptation (see explanations of the terms "quick speed adaptation of ACC" and "slow speed adaptation of ACC" in Sect. 23.6 of the book [33]), which are randomly distributed in traffic flow of usual vehicles without ACC, no considerable change occurs in traffic flow: Traffic breakdown occurs after a time delay at the on-ramp bottleneck (Fig. 6(c, d)).

The situation changes qualitatively, if we further increase the percentage of the ACC vehicles: At 20 % of ACC vehicles there is no traffic breakdown during the observation time (Fig. 6(e, f)).

However, if we increase the time of observation of traffic flow, we find that at 20 % of ACC vehicles traffic breakdown can nevertheless occur with a considerably longer mean time delay (Fig. 7(a, b)). If we increase the percentage of the ACC vehicles up to 30 %, no traffic breakdown occurs at the bottleneck any more (Fig. 7(c, d)).

Thus, we can conclude that when the percentage of ACC vehicles with quick speed adaptation increases in traffic flow, firstly the mean time delay in traffic breakdown increases (Fig. 7(a, b)). A further increase in the percentage of the ACC vehicle prevents traffic breakdown at a highway bottleneck (Fig. 7(c, d)).

11. Conclusions

1. The fundamental empirical basis for automatic driving as well as for reliable control and optimization of vehicular traffic and transportation networks is the set of empirical features of traffic breakdown at a road bottleneck.
2. The theoretical fundament resulting from three-phase traffic theory for the development of reliable control and optimization of traffic and transportation networks is the existence of the range of the infinite number of highway capacities: At any time instant, there are the infinite number of highway capacities within a range of the flow rate between the minimum capacity and the maximum capacity (Fig. 4); within this flow range, traffic breakdown can be induced at the bottleneck.
3. The explanation of traffic breakdown at a highway bottleneck by an F → S transition in a metastable free flow introduced in three-phase traffic theory is responsible for the incommensurability of three-phase traffic theory with all other traffic flow theories.
4. Classical traffic theories failed to explain the set of empirical features of traffic breakdown at a highway bottleneck. For this reasons, traffic flow models, which are based on these classical traffic theories, cannot be used for a reliable analysis of the impact of automatic driving and/or other ITS-applications on traffic flow. Examples of these traffic flow models are the classical LWR-model, Daganzo's Cell Transmission Model, GM car-following model of Herman, Gazis et al., Newell optimal velocity (OV) model, Gipps model, Bando et al. OV model, Payne's macroscopic model, Aw-Rasche macroscopic model, Treiber's IDM, Nagel-Schreckenberg cellular automaton (CA) model as well as well-known and generally used traffic simulation tools (like VISSIM (Wiedemann-model) and SUMO (Krauß-model)): Simulations of the effect of automatic driving and/or other ITS-applications on traffic flow with the use of such traffic simulation models and tools lead to incorrect results and invalid conclusions.
5. To perform reliable analysis of the impact of automatic driving and/or other ITS-applications on traffic flow, traffic flow models in the framework of three-phase traffic theory should be used. This is because these models can explain the set of empirical features of traffic breakdown at a road bottleneck.

Acknowledgements

We thank our partners for their support in the project "UR:BAN—Urban Space: User oriented assistance systems and network management", funded by the German Federal Ministry of Economics and Technology. I thank Sergey Klenov for discussions and help in simulations.

References

1. Kerner, B. S. (2013): Criticism of generally accepted fundamentals and methodologies of traffic and transportation theory: a brief review. Physica A, 392, 5261–5282.
2. Lighthill, M. J., Whitham, G. B. (1955): On kinematic waves: theory of traffic flow on long crowded roads. Proc. R. Soc. A, Math. Phys. Eng. Sci., 229, 281–345.
3. Richards, P. I. (1956): Shockwaves on the highway. Oper. Res., 4, 42–51.
4. Daganzo, C. F. (1994): The cell transmission model: a simple dynamic representation of highway traffic. Transp. Res., Part E, Logist. Transp. Rev., 28, 269–287.
5. Daganzo, C. F. (1995): The cell transmission model, part II: network traffic. Transp. Res., Part B, Methodol., 29, 79–93.
6. Chandler, R. E., Herman, R., Montroll, E. W. (1958): Traffic dynamics: studies in car following. Oper. Res., 6, 165–184.
7. Herman, R., Montroll, E. W., Potts, R. B., Rothery, R. W. (1959): Traffic dynamics: analysis of stability in car following. Oper. Res., 7, 86–106.
8. Gazis, D. C., Herman, R., Rothery, R. W. (1961): Nonlinear follow-the-leader models of traffic flow. Oper. Res., 9, 545–567.
9. Gazis, D. C. (2002): Traffic theory. Berlin: Springer.
10. Kerner, B. S., Konhäuser, P. (1993): Cluster effect in initially homogeneous traffic flow. Phys. Rev. E, 48, 2335–2338.
11. Kerner, B. S., Konhäuser, P. (1994): Structure and parameters of clusters in traffic flow. Phys. Rev. E, 50, 54–83.
12. May, A. D. (1990): Traffic flow fundamentals. New Jersey: Prentice-Hall.
13. Chowdhury, D., Santen, L., Schadschneider, A. (2000): Statistical physics of vehicular traffic and some related systems. Phys. Rep., 329, 199.
14. Helbing, D. (2001): Traffic and related self-driven many-particle systems. Rev. Mod. Phys., 73, 1067–1141.
15. Highway capacity manual (2000): National Research Council, Transportation Research Board, Washington, D.C.
16. Highway capacity manual (2010): National Research Council, Transportation Research Board, Washington, D.C.
17. Nagatani, T. (2002): The physics of traffic jams. Rep. Prog. Phys., 65, 1331–1386.
18. Nagel, K., Wagner, P., Woesler, R. (2003): Still flowing: approaches to traffic flow and traffic jam modeling. Oper. Res., 51, 681–716.
19. Treiber, M., Kesting, A. (2013): Traffic flow dynamics. Heidelberg: Springer.
20. Elefteriadou, L. (2014): An introduction to traffic flow theory. Springer optimization and its applications (Vol. 84). Berlin: Springer.
21. Greenshields, B. D. (1935): A study of traffic capacity. Proceedings of the Annual Meeting - Highway Research Board, 14, 448–477.
22. Elefteriadou, L., Roess, R. P., McShane, W. R. (1995): Probabilistic nature of breakdown at freeway merge junctions. Transp. Res. Rec., 1484, 80–89.
23. Hall, F. L., Agyemang-Duah, K. (1991): Freeway capacity drop and the definition of capacity. Transp. Res. Rec., 1320, 91–98.
24. Hall, F. L., Hurdle, V. F., Banks, J. H. (1992): Synthesis of recent work on the nature of speed-flow and flow-occupancy (or density) relationships on freeways. Transp. Res. Rec., 1365, 12–18.
25. Persaud, B., Yagar, S., Brownlee, R. (1998): Exploration of the breakdown phenomenon in freeway traffic. Transp. Res. Rec., 1634, 64–69.
26. Lorenz, M., Elefteriadou, L. (2000): A probabilistic approach to defining freeway capacity and breakdown. Trans. Res. Circ., E-C018, 84–95.
27. Brilon, W. (2004): Zufallscharakter der Kapazität von Autobahnen und praktische Konsequenzen. In Deutscher straßen- und Verkehrskongress (Berichtsband), Berlin, 13.–15. Okt. 2004, ISBN 3-937356-67-3.
28. Brilon, W., Regler, M., Geistefeld, J. (2005): Zufallscharakter der Kapazität von Autobahnen und praktische Konsequenzen. Straßenverkehrstechnik, 3, 136.
29. Brilon, W., Geistefeldt, J., Regler, M. (2005): Reliability of freeway traffic flow: a stochastic concept of capacity. In H. S. Mahmassani (Ed.), Traffic and transportation theory (pp. 125–144). Amsterdam: Elsevier Science.
30. Wardrop, J. G. (1952): Some theoretical aspects of road traffic research. Proc. Inst. Civil Eng., 1, 325–362.
31. Bell, M. G. H., Iida, Y. (1997): Transportation network analysis. Chichester: Wiley.
32. Kerner, B. S. (2009): Introduction to modern traffic flow theory and control. Heidelberg: Springer.
33. Kerner, B. S. (2004): The physics of traffic. Berlin: Springer.
34. Kerner, B. S., Koller, M., Klenov, S. L., Rehborn, H., Leibel, M. (2015): The physics of empirical nuclei for spontaneous traffic breakdown in free flow at highway bottlenecks. Physica A, 438, 365–397.
35. Kerner, B. S. (1999): Congested traffic flow: observations and theory. Transp. Res. Rec., 1678, 160–167.
36. Kerner, B. S. (1999): The physics of traffic. Phys. World, 12, 25–30.
37. Kerner, B. S. (1999): Theory of congested traffic flow: self-organization without bottlenecks. In A. Ceder (Ed.), Transportation and traffic theory (pp. 147–171). Amsterdam: Elsevier Science.
38. Kerner, B. S., Klenov, S. L., Schreckenberg, M. (2011): Simple cellular automaton model for traffic breakdown, highway capacity, and synchronized flow. Phys. Rev. E, 84, 046110.
39. Kerner, B. S., Klenov, S. L., Wolf, D. E. (2002): Cellular automata approach to three-phase traffic theory. J. Phys. A, Math. Gen., 35, 9971–10013.
40. Kerner, B. S. (2011): Physics of traffic gridlock in a city. Phys. Rev. E, 84, 045102(R).
41. Kerner, B. S. (2013): The physics of green-wave breakdown in a city. Europhys. Lett., 102, 28010.
42. Kerner, B. S. (2014): Three-phase theory of city traffic: moving synchronized flow patterns in under-saturated city traffic at signals. Physica A, 397, 76–110.
43. Kerner, B. S., Klenov, S. L., Schreckenberg, M. (2014): Traffic breakdown at a signal: classical theory versus the three-phase theory of city traffic. J. Stat. Mech.: Theory Exp., 2014, P03001.
44. Kerner, B. S., Klenov, S. L., Hermanns, G., Hemmerle, P., Rehborn, H., Schreckenberg, M. (2013): Synchronized flow in oversaturated city traffic. Phys. Rev. E, 88, 054801.
45. Kerner, B. S., Hemmerle, P., Koller, M., Hermanns, G., Klenov, S. L., Rehborn, H., Schreckenberg, M. (2014): Empirical synchronized flow in oversaturated city traffic. Phys. Rev. E, Stat. Nonlinear Soft Matter Phys., 90, 032810.
46. Kuhn, T. S. (2012): The structure of scientific revolutions. 4th ed. Chicago: The University of Chicago Press.
47. Kerner, B. S., Klenov, S. L., Schreckenberg, M. (2014): Probabilistic physical characteristics of phase transitions at highway bottlenecks: incommensurability of three-phase and two-phase traffic-flow theories. Phys. Rev. E, 89, 052807.
48. Hoogendoorn, S. P., Knoop, V. L., Van Zuylen, H. J. (2008): Robust control of traffic networks under uncertain conditions. J. Adv. Transp., 42, 357–377.
49. Zhang, C., Chen, X., Sumalee, A. (2011): Robust wardrop's user equilibrium assignment under stochastic demand and supply: expected residual minimization approach. Transp. Res., Part B, Methodol., 45, 534–552.
50. Kerner, B. S. (2011): Optimum principle for a vehicular traffic network: minimum probability of congestion. J. Phys. A, Math. Theor., 44, 092001.
51. Kerner, B. S. (2007): On-ramp metering based on three-phase traffic theory. Traffic Eng. Control, 48(1), 28–35; (2), 68–75; (3), 114–120.
52. Kerner, B. S. (2005): Control of spatiotemporal congested traffic patterns at highway bottlenecks. Physica A, 355, 565–601.
53. Kerner, B. S. (2007): Control of spatiotemporal congested patterns at highway bottlenecks. IEEE Trans. Intell. Transp. Syst., 8, 308–320.
54. Kerner, B. S. (2008): On-ramp metering based on three-phase traffic theory: downstream off-ramp and upstream on-ramp bottlenecks. Transp. Res. Rec., 2088, 80–89.
55. Kerner, B. S. (1998): Experimental properties of self-organization in traffic flow. Phys. Rev. Lett., 81, 3797–3800.
56. Kerner, B. S. (2001): Complexity of synchronized flow and related problems for basic assumptions of traffic flow theories. Netw. Spat. Econ., 1, 35–76.
57. Kerner, B. S., Rehborn, H., Schäfer, R.-P., Klenov, S. L., Palmer, J., Lorkowski, S., Witte, N. (2013): Traffic dynamics in empirical probe vehicle data studied with three-phase theory: spatiotemporal reconstruction of traffic phases and generation of jam warning messages. Physica A, 392, 221–251.
58. Ktousakis, I. A., Nokolos, I. K., Papageorgiou, M. (2015): On microscopic modeling of adaptive cruise control systems. Transp. Res. Proc., 9, 111–127.
59. Shladover, S. E., Su, D., Lu, X.-Y. (2012): Impacts of cooperative adaptive cruise control on freeway traffic flow. Transp. Res. Rec., 2324, 63–70.
60. Wagner, P. (2015): Steuerung und Management in einem Verkehrssystem mit autonomen Fahrzeugen. In M. Maurer, J. Ch. Gerdes, B. Lenz, H. Winner (Eds.), Autonomes Fahren (pp. 313–330). Berlin: Springer.
61. Friedrich, B. (2015): Verkehrliche Wirkung autonomer Fahrzeuge. In M. Maurer, J. Ch. Gerdes, B. Lenz, H. Winner (Eds.), Autonomes Fahren (pp. 331–350). Berlin: Springer.
62. Roncoli, C., Papageorgiou, M., Papamichail, I. (2015): Traffic flow optimization in presence of vehicle automation and communication systems, part I: a first-order multi-lane model for motorway traffic. Transp. Res., Part C, Emerg. Technol., 57, 241–259.
63. Motamedidehkordi, N., Benz, T., Margreiter, M. (2016): Shockwave analysis on motorways and possibility of damping by autonomous vehicles. In Advanced microsystems for automotive applications 2015. Lecture notes in mobility (Vol. 2016, pp. 37–52). Berlin: Springer.
64. Kerner, B. S., Klenov, S. L. (2002): A microscopic model for phase transitions in traffic flow. J. Phys. A, Math. Gen., 35, L31–L43.
65. Kerner, B. S., Klenov, S. L. (2003): Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks. Phys. Rev. E, Stat. Nonlinear Soft Matter Phys., 68, 036130.
66. Kerner, B. S., Klenov, S. L. (2009): Phase transitions in traffic flow on multilane roads. Phys. Rev. E, Stat. Nonlinear Soft Matter Phys., 80, 056101.
67. Gardiner, C. W. (1990): Handbook of stochastic methods. 2nd ed. Berlin: Springer.

68. Kerner, B. S. (2015): Test bed for simulations of the effect of a vehicle ad hoc network on traffic flow. In W. Chen (Ed.), Vehicular communications and networks (pp. 223–254). Cambridge: Woodhead Publishing.
69. Kerner, B. S., Klenov, S. L. (2009): Traffic breakdown: probabilistic theory of. In R. A. Meyers (Ed.), Encyclopedia of complexity and system science (pp. 9282–9302). Berlin: Springer.
70. Kerner, B. S., Klenov, S. L. (2006): Probabilistic breakdown phenomenon at on-ramp bottlenecks in three-phase traffic theory: congestion nucleation in spatially non-homogeneous traffic. Physica A, 364, 473–492.
71. Kerner, B. S., Klenov, S. L. (2006): Probabilistic breakdown phenomenon at on-ramp bottlenecks in three-phase traffic theory. Transp. Res. Rec., 1965, 70–78.

Appendix A: The classical understanding of the nature of stochastic highway capacity

To illustrate the above critical conclusion about the generally accepted fundamentals and methodologies of traffic and transportation theory (Sects. 2 and 5), we discuss here the criticism of the generally accepted understanding of stochastic highway capacity of free flow at a highway bottleneck [1, 32].

The classical highway capacity is defined through the occurrence of traffic breakdown at a bottleneck: The highway capacity is equal to the flow rate in an initially free flow at the bottleneck at which traffic breakdown is observed at the bottleneck (see e.g., [12, 15, 16, 20–29]).

During last 20 years it was found that empirical traffic breakdown exhibits a probabilistic character and the probability of the spontaneous breakdown is an increasing flow rate function $P^{(B)}(q_{\rm sum})$, where $q_{\rm sum}$ is the flow rate in an initially free flow at the bottleneck (see references in the book [20]). Respectively, Brilon has introduced the following concept for stochastic highway capacity [27–29].

In accordance with the classical capacity definition, Brilon's stochastic highway capacity C is equal to the flow rate $q_{\rm sum}$ at the bottleneck. At any time instant, there is a particular value of stochastic capacity of free flow at the bottleneck.

However, as long as free flow is observed at the bottleneck, this particular value of stochastic capacity cannot be measured. Therefore, stochastic capacity is defined through a capacity distribution function $F_C^{(B)}(q_{\rm sum})$ [27–29]:

$$F_C^{(B)}(q_{\rm sum}) = p(C \leq q_{\rm sum}), \quad (8)$$

where $p(C \leq q_{\rm sum})$ is the probability that stochastic highway capacity C is equal to or smaller than the flow rate $q_{\rm sum}$ in free flow at the bottleneck.

Thus the basic theoretical assumption of the classical understanding of stochastic highway capacity is that traffic breakdown is observed at a time instant t at which the flow rate $q_{\rm sum}(t)$ reaches the capacity $C(t)$. This means that the flow rate function of the probability of traffic breakdown $P^{(B)}(q_{\rm sum})$ should be determined by the capacity distribution function $F_C^{(B)}(q_{\rm sum})$ [27–29]:

$$P^{(B)}(q_{\rm sum}) = F_C^{(B)}(q_{\rm sum}). \quad (9)$$

It must be noted that the breakdown probability function $P^{(B)}(q_{\rm sum})$ found in empirical observations [20, 25–29] is the *empirical evidence*. However, condition (9) is a *theoretical hypothesis* only. This is because in contrast with the breakdown probability function $P^{(B)}(q_{\rm sum})$, the capacity distribution function $F_C^{(B)}(q_{\rm sum})$ cannot be measured. Below we explain why the hypothesis (9) and, therefore, Brilon's stochastic capacity contradicts the set of fundamental empirical features of traffic breakdown.

This understanding of stochastic capacity of free flow at a bottleneck, which is currently well accepted in the community of traffic and transportation researchers [20], is illustrated in Fig. 8.

In Fig. 8, we show a qualitative hypothetical fragment of the time-dependence of stochastic capacity $C(t)$ over time t. Left in Fig. 8, a qualitative flow rate dependence of the probability of spontaneous traffic breakdown $P^{(B)}(q_{\rm sum})$ is shown that is the same as that in Fig. 4(a). In accordance with Eq. (9), capacity $C(t)$ can stochastically change over time (Fig. 8).

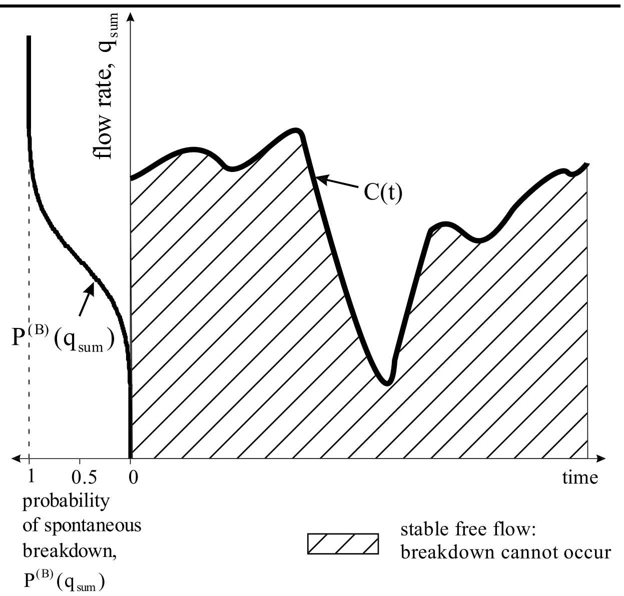

Fig. 8. Qualitative explanation of Brilon's stochastic highway capacity of free flow at a highway bottleneck. The flow rate function of the probability of the spontaneous breakdown $P^{(B)}(q_{\rm sum})$ (*left figure*) is the same as that shown in Fig. 4(a)

It is often assumed that a stochastic behavior of highway capacity is associated with a stochastic change in traffic parameters over time [20, 27–29]. Examples of the traffic parameters, which can indeed be stochastic time-functions in real traffic, are weather, mean driver's characteristics (e.g., mean driver reaction time), share of long vehicles, etc.

In accordance with the definition of stochastic capacity (8), (9), no traffic breakdown can occur, when the time dependence of the flow rate is given by a hypothetical time dependence $q_{\rm sum}(t) = q^{(1)}_{\rm sum}(t)$. This is because at all-time instants $q^{(1)}_{\rm sum}(t) < C(t)$ (Fig. 9(a)).

In contrast, for another hypothetical time dependence $q_{\rm sum}(t) = q^{(2)}_{\rm sum}(t)$ traffic breakdown should occur at time instant t_1 at which $q^{(2)}_{\rm sum}(t_1) = C(t_1)$, i.e., this flow rate is equal to the capacity value (Fig. 9(b)).

In other words, the classical understanding of a particular value of stochastic capacity can be explained as follows: At a *given time instant* no traffic breakdown can occur at a highway bottleneck if the flow rate in free flow at the bottleneck at the time instant is smaller than the value of the capacity *at this time instant*.

The basic importance of the words "at a given time instant" in the capacity definition is as follows: Brilon's stochastic capacity $C(t)$ changes stochastically over time (Fig. 8). Thus at a given time instant traffic breakdown can occur at the flow rate that is smaller than the value of the stochastic capacity was at another time instant.

In the classical understanding of stochastic capacity, free flow is *stable* under condition $q_{\rm sum}(t) < C(t)$. This means that *no* traffic

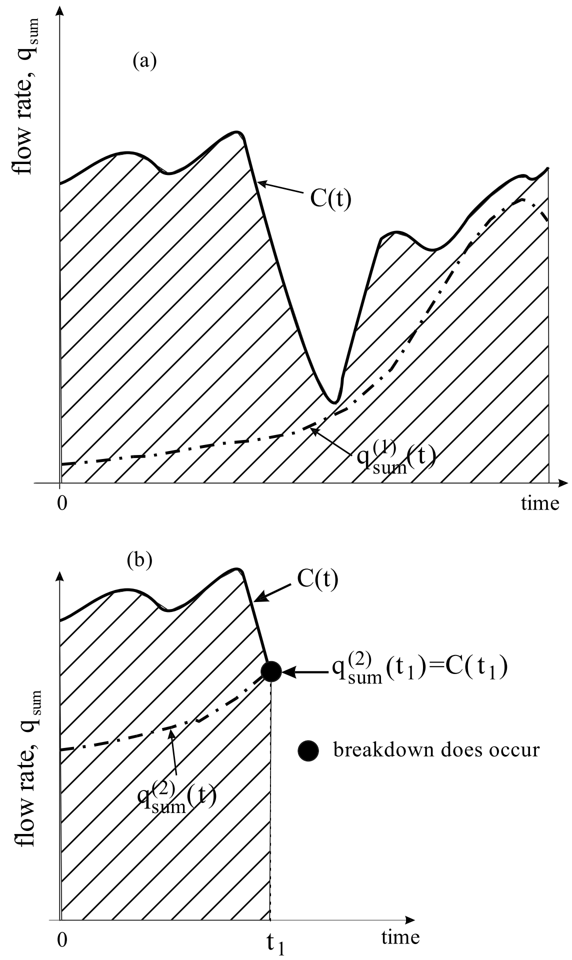

Fig. 9. Qualitative explanation of traffic breakdown with the use of Brilon's stochastic highway capacity of free flow at a highway bottleneck. The fragment of the hypothetical time-function of Brilon's stochastic highway capacity *C(t)* is taken from Fig. 8

breakdown can occur or be induced at the bottleneck at long as the flow rate in free flow at the bottleneck is smaller than the stochastic capacity. This contradicts to the empirical fact that traffic breakdown can be induced at the bottleneck due to the upstream propagation of a localized congested pattern (Fig. 2(b)).

This is because stochastic highway capacity *cannot* depend on whether there is a congested pattern, which has occurred outside of the bottleneck and independent of the bottleneck existence, or not. Indeed, the empirical evidence of induced traffic breakdown is the empirical proof that at a given flow rate at a bottleneck there can be *one of two* different traffic states at the bottleneck: (i) A traffic state related to free flow and (ii) a congested traffic state labeled as synchronized flow in Fig. 2(b). Due to the upstream propagation of a localized congested pattern, a transition from the state of free flow to the state of synchronized flow, i.e., traffic breakdown is induced (see Appendix B).

The induced traffic breakdown is impossible to occur under the classical understanding of the nature of highway capacity [12, 20, 27–29]. This is because in this classical understanding of highway capacity, free flow is *stable* under condition $q_{sum}(t) < C(t)$, i.e., no traffic breakdown can occur (Fig. 9(a)).

In contrast with this classical understanding of the nature of highway capacity, the evidence of the empirical induced breakdown means that free flow is in a *metastable* state with respect to the breakdown. The metastability of free flow at the bottleneck should exist for all flow rates at which traffic breakdown *can be induced* at the bottleneck. This empirical evidence of the metastability of free flow at the bottleneck contradicts fundamentally the concept of Brilon's stochastic capacity, in which free flow is *stable* under condition $q_{sum}(t) < C(t)$.

Thus the currently accepted understanding of stochastic highway capacity [20, 27–29] failed because this understanding about the nature of highway capacity contradicts the empirical evidence that traffic breakdown can be induced at a highway bottleneck as observed in real traffic (Fig. 2(b)) (see also Appendix B).

Appendix B: Empirical induced breakdown—empirical proof of the metastability of free flow at highway bottlenecks

In this appendix, following empirical studies of traffic breakdown presented in [1, 32–34], we show the special importance of the following two empirical features of traffic breakdown at a highway bottleneck:

i. The downstream front of congested traffic resulting from the breakdown is usually fixed at the bottleneck location; as abovementioned this congested traffic is called synchronized flow (S) (Fig. 2).
ii. Empirical observations of induced traffic breakdown at highway bottlenecks (Fig. 2(b)).

The importance of these empirical features of traffic breakdown is as follows: They prove that traffic breakdown is an $F \rightarrow S$ transition occurring in a metastable free flow at a highway bottleneck [33–37]. In more details, the empirical prove of the metastability of free flow at a highway bottleneck one can find in [34].

However, before we consider the empirical proof of the metastability of free flow at highway bottlenecks, we should define and explain the term "nucleus" for traffic breakdown (Appendix B.1) as well as define "empirical spontaneous traffic breakdown" and "empirical induced traffic breakdown" (Appendix B.2).

B.1. Explanation of nucleus for traffic breakdown

The term "metastable free flow with respect to an $F \rightarrow S$ transition" means that a small enough disturbance for free flow at a bottleneck decays; therefore, in this case free flow persists at the bottleneck over time. However, when a critical disturbance (or a disturbance that is larger than the critical one) appears in free flow in a neighborhood of the bottleneck, traffic breakdown occurs at the bottleneck. In accordance with general theory of metastable systems of natural science [67], such a (speed, density and/or flow rate) disturbance in free traffic flow can be called a *nucleus* for traffic breakdown ($F \rightarrow S$ transition) at a bottleneck.

For this reason, the term "the metastability of free flow with respect to the $F \rightarrow S$ transition" means that traffic breakdown at the bottleneck exhibits the *nucleation nature*: If the nucleus for traffic breakdown occurs in free flow at the bottleneck, traffic breakdown does occur. In contrast, as long as no nucleus appears, no breakdown occurs in a metastable state of free flow. It must be noted that there are two ways for nucleus occurrence:

1. The nucleus for traffic breakdown can occur *spontaneously* in free flow, for example, through random fluctuations of the free flow speed, the density, or/and the flow rate at the bottleneck.

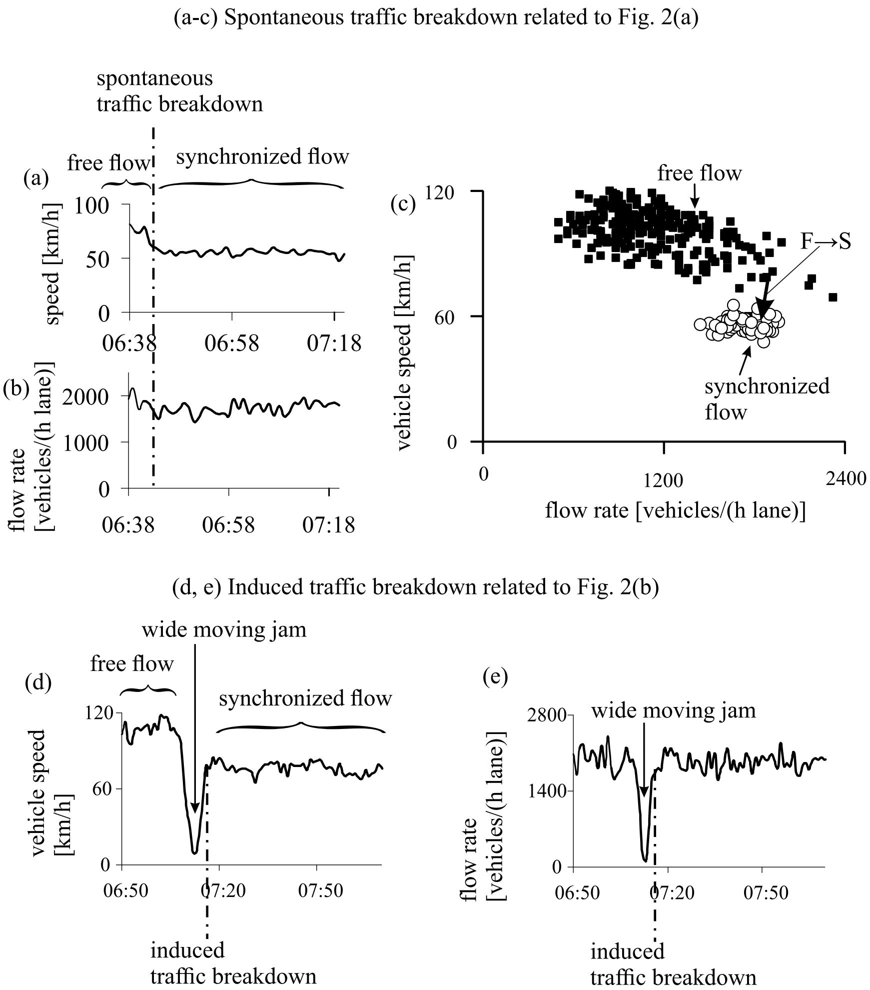

Fig. 10. Empirical features of synchronized flow occurring through spontaneous traffic breakdown (a–c) and induced traffic breakdown (d, e) related to Figs. 2(a) and 2(b), respectively (1 min averaged data). (a, b, d, e) Average speed (a, d) and flow rate (b, e) as time-function related to locations 6.4 km (a, b) and 16.2 km (d, e) associated with merging regions of on-ramp bottlenecks on freeway A5-South in Germany (see schema of freeway section in Fig. 2.1 of the book [33]). (c) Data related to (a, b) in the speed-flow rate plane; *arrow* F → S marks spontaneous traffic breakdown whose duration is about 1 min

Recently empirical nuclei for spontaneous traffic breakdown have been revealed in studied of real field traffic data [34].

2. The nucleus for traffic breakdown can be *induced* in free flow at the bottleneck. There can be the following scenario for the induced traffic breakdown in *real* free flow at the bottleneck. Firstly, a local congested pattern occurs at a downstream bottleneck. Then the pattern propagates upstream to the location of the bottleneck under consideration. When this congested pattern reaches the bottleneck, the pattern induces traffic breakdown at the bottleneck. In accordance with above consideration of the nucleation nature of traffic breakdown, this local congested pattern can be considered the nucleus that induces traffic breakdown at the bottleneck.

B.2. Definitions of empirical spontaneous and induced traffic breakdowns

This consideration can explain why in *empirical data* (i.e., field data measured in real traffic) we distinguish two typed of traffic breakdowns at highway bottlenecks: (i) Empirical spontaneous traffic breakdown. (ii) Empirical induced traffic breakdown [32–34].

- *Empirical spontaneous traffic breakdown* is defined as follows. If before traffic breakdown occurs at the bottleneck, there is free flow at the bottleneck as well as upstream and downstream in a neighborhood of the bottleneck, then traffic breakdown at the bottleneck is called spontaneous traffic breakdown (Fig. 2(a)).
- *Empirical induced traffic breakdown* at the bottleneck is traffic breakdown induced by the propagation of a spatiotemporal congested traffic pattern. This congested pattern has occurred earlier than the time instant of traffic breakdown at the bottleneck and at a different road location (for example at a downstream bottleneck) than the bottleneck location (Fig. 2(b)).

Example of empirical spontaneous traffic breakdown is shown in Figs. 2(a) and 10(a–c). This is a well-known traffic breakdown studied by many researchers (see, e.g., [12, 15, 16, 20, 22–29] and references there). It is also well-known that states of free flow at a bottleneck overlap in the flow rate with states of synchronized flow (congested traffic) measured at the bottleneck location (Fig. 10). However, this well-known empirical fact considered as a solely empirical fact does *not* prove that real free flow is a *metastable* state with an F → S transition.

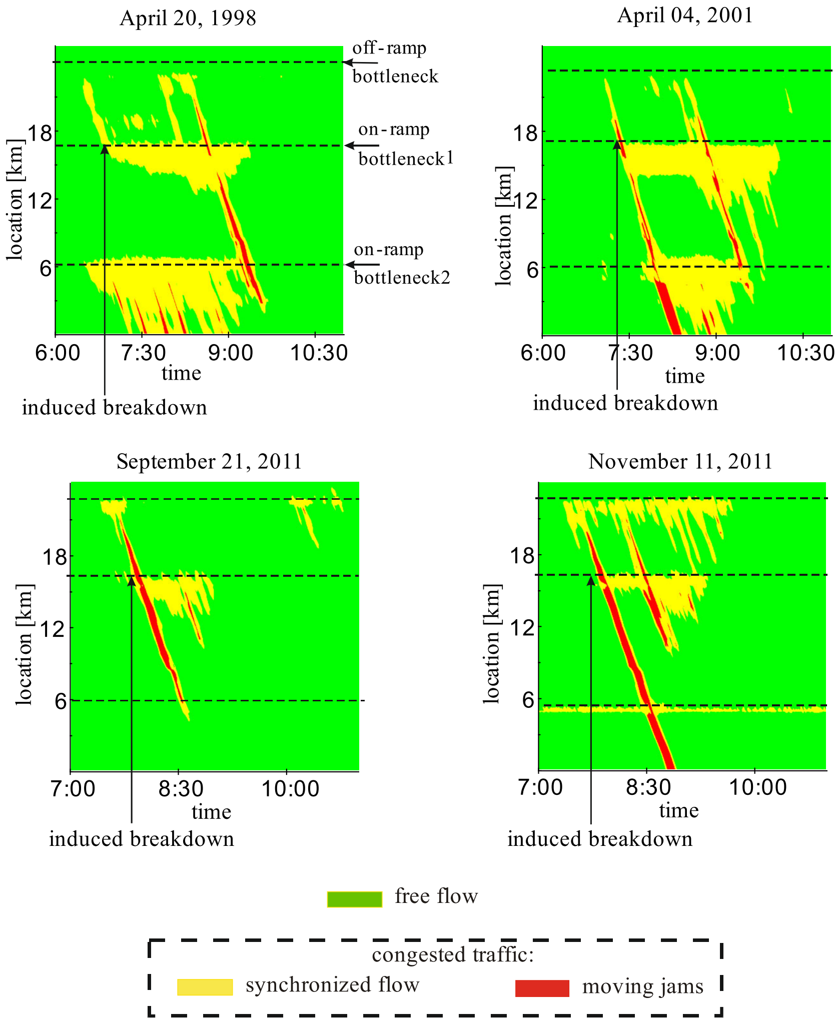

Fig. 11. Examples of empirical induced traffic breakdown at highway bottlenecks measured on four different days by road detectors installed on freeway A5-South in Germany (1 min averaged field data): Representation of measured speed data in the time–space plane; explanations of the reconstruction of spatiotemporal traffic dynamics based on road detector measurements shown in these figures are given in Sect. C.2 of [57]. Freeway section schema is shown in Fig. 2.1 of the book [33]

The proof of the metastability of free flow with respect to traffic breakdown (F → S transition) is the empirical evidence of induced F → S transition caused by the propagation of a local spatiotemporal congested traffic pattern (Fig. 2(b)) [32–34]. Thus the local congested pattern is the nucleus for the transition between these two flow phases F and S.

After the induced breakdown has occurred, the emergent synchronized flow can persist over time at the bottleneck independent on a further behavior of the congested pattern that has initially induced the breakdown. In the case shown in Fig. 2(b), this pattern is a wide moving jam that propagates far away from the bottleneck, while the induced synchronized flow remains to be localized at the bottleneck. Synchronized flow that results from this induced synchronized flow at the bottleneck (Fig. 10(d, e)) exhibits qualitatively the same empirical features as those of synchronized flow occurring through empirical spontaneous breakdown. A more detailed consideration of empirical features of spontaneous and induced traffic breakdowns can be found in [34].

There can be many different scenarios of empirical induced traffic breakdowns (Fig. 11). All these scenarios show qualitatively the same nucleation nature of traffic breakdown at highway bottlenecks discussed in this appendix. Therefore, we can make the conclusion:

- Empirical induced breakdown is the empirical proof of the metastability of free flow at highway bottlenecks.

It should be noted that a number of other empirical examples of traffic breakdown at highway bottlenecks that substantiate the above conclusion can be found in Chap. 2 and Part II of the book [33], in Chaps. 2, 3, 5, and 7 of the book [32] as well as in [34].

B.3. Empirical induced traffic breakdown as one of the consequences of spillover

Most of the traffic researchers (see, e.g., [12, 15, 16, 20, 22–29] and references there) do not consider the empirical evidence of empirical induced traffic breakdown. The upstream propagation of traffic congestion occurring at a downstream bottleneck is usually called by traffic researchers as *spillback*. If this traffic congestion forces congested traffic at an upstream bottleneck, it is called the *spillover* effect.

When the wide moving jam shown in Fig. 2(b) reaches the bottleneck, the jam can indeed be considered spillover: The jam forces congested traffic at the bottleneck. However, due to the upstream jam propagation, the jam can be considered as spillover only during a short time interval: When the jam is far away upstream of the bottleneck, the jam does not force congested traffic at the bottleneck any more.

In [34], we have explained the reason why we do not use the term *spillover*: This is because there can be at least the following qualitatively different empirical effects of spillover:

(i) An empirical induced traffic breakdown occurs due to congested pattern propagation through a bottleneck (Figs. 2(b) and 11).

(ii) The jam propagation through a bottleneck does not lead to induced traffic breakdown (see Fig. 16(b) of [34]).

(iii) An expanded congested pattern (EP) occurs due to spillover at a bottleneck (see Fig. 18(b) of [34]). This spillover *cannot* be considered as induced traffic breakdown. This is because during the whole time of the existence of traffic congestion at the bottleneck this traffic congestion is forced by downstream traffic congestion.

Therefore, rather than consider all these qualitatively different traffic phenomena as the same effect *spillover*, to understand real vehicular traffic, one should consider each of these cases of spillover separately each other, i.e., as qualitatively different traffic phenomena.

Appendix C: Infinite number of stochastic highway capacities of three-phase traffic theory

Follow [68], in this Appendix we explain the understanding of stochastic capacity introduced in the three-phase traffic theory [32, 33]. As explained in Appendix A, in the classical understanding of stochastic capacity, condition (9) is assumed. However, the condition (9) contradicts the empirical fact about observations of induced traffic breakdowns (Appendix B).

C.1. Basic assumption of three-phase traffic theory about the nature of traffic breakdown at highway bottlenecks

In contrast with condition (9), in three-phase traffic theory the following basic assumption about the nature of traffic breakdown (F → S transition) is made [32, 33, 35–37]:

$$P^{(B)}(q_{\mathrm{sum}}) = P^{(B)}_{\mathrm{nucleus}}(q_{\mathrm{sum}}), \qquad (10)$$

where $P^{(B)}_{\mathrm{nucleus}}(q_{\mathrm{sum}})$ is the flow-rate dependence of the probability that during a given time interval (that is the same as that used in the definition of the breakdown probability $P^{(B)}(q_{\mathrm{sum}})$) a nucleus for traffic breakdown occurs spontaneously in free flow at a bottleneck. A related mathematical nucleation theory of traffic breakdown can be found in [69–71].

For qualitative explanations of condition (10), firstly we assume that traffic parameters (weather, mean driver's characteristics, share of long vehicles, etc,) remain the same for all flow rates in free flow. Under this condition, we can also assume that the larger is the flow rate q_{sum} in free flow at the bottleneck, the smaller is the nucleus required for the breakdown at a bottleneck. Obviously, the probability of the occurrence of a small speed disturbance in free flow is considerably larger than the probability of the occurrence of a large disturbance. This means that probability of the spontaneous occurrence of a nucleus for traffic breakdown $P^{(B)}_{\mathrm{nucleus}}(q_{\mathrm{sum}})$ is an increasing function of the flow rate q_{sum}. In accordance with (10), this explains the increasing flow rate function of the breakdown probability $P^{(B)}(q_{\mathrm{sum}})$ (Fig. 4(a)).

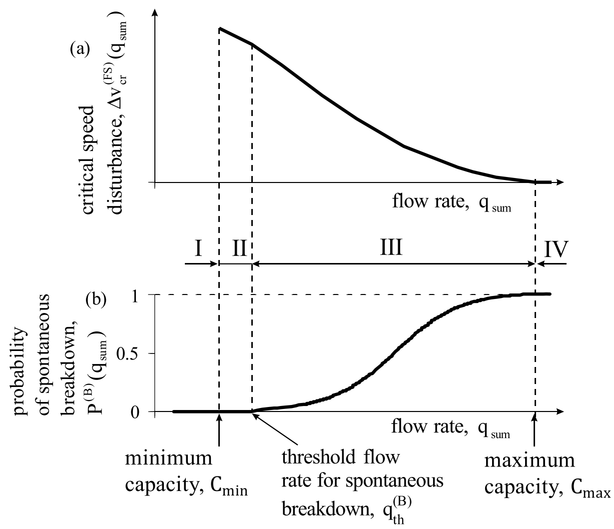

Fig. 12. Explanation of condition (10): (a) Qualitative flow rate dependence of function $\Delta v^{(FS)}_{cr}(q_{\mathrm{sum}})$. (b) Breakdown probability $P^{(B)}(q_{\mathrm{sum}})$ taken from Fig. 4(a). Flow rate ranges I, II, III, and IV have the same sense as those shown in Fig. 4(a)

As an example of this general discussion of condition (10), we consider the occurrence of a nucleus associated with a time-limited critical local decrease in the speed in an initial free flow at a bottleneck denoted by $\Delta v^{(FS)}_{cr}$ (Fig. 12(a)). The larger the flow rate q_{sum} in free flow at the bottleneck, the smaller should be the value $\Delta v^{(FS)}_{cr}(q_{\mathrm{sum}})$ that initiates traffic breakdown at the bottleneck. The related decreasing function $\Delta v^{(FS)}_{cr}(q_{\mathrm{sum}})$, which is qualitatively shown in Fig. 12(a), has indeed been found in simulations with Kerner-Klenov stochastic microscopic three-phase traffic flow model [64].

Condition (10) explains flow rate ranges II–IV discussed in Sect. 6 of the main text as follows (Fig. 12). In flow rate range II (condition (4)), a very large value $\Delta v^{(FS)}_{cr}(q_{\mathrm{sum}})$ (large nucleus) is required for the breakdown, so we can assume that the probability of spontaneous occurrence of such very large speed disturbance in free flow during a given time interval is zero, i.e., $P^{(B)}_{\mathrm{nucleus}}(q_{\mathrm{sum}}) = 0$. In accordance with (10), the probability of spontaneous breakdown $P^{(B)}(q_{\mathrm{sum}}) = 0$. This means that in this case only induced traffic breakdown is possible.

In flow rate range III (condition (5)), the value $\Delta v^{(FS)}_{cr}(q_{\mathrm{sum}})$ required for the breakdown decreases sharply. Therefore, the probability of the spontaneous occurrence of such a speed disturbance due to fluctuations in free flow during a given time interval can satisfy conditions $0 < P^{(B)}_{\mathrm{nucleus}}(q_{\mathrm{sum}}) < 1$.

In flow rate range IV (condition (6)), the value $\Delta v^{(FS)}_{cr}(q_{\mathrm{sum}})$ required for the breakdown is as small as zero; therefore, the probability of the spontaneous occurrence of a nucleus for traffic breakdown $P^{(B)}_{\mathrm{nucleus}}(q_{\mathrm{sum}}) = 1$. Therefore, in accordance with (10), the probability of spontaneous traffic breakdown $P^{(B)}(q_{\mathrm{sum}}) = 1$.

C.2. Maximum and minimum capacities as stochastic functions

It must be noted that the maximum capacity C_{max}, the minimum capacity C_{min}, and the value $q^{(B)}_{\mathrm{th}}$ depend on traffic parameters, like weather, mean driver's characteristics (e.g., mean driver reaction time), share of long vehicles, etc. In real traffic flow, these traffic parameters change over time. For this reason, the values C_{max}, C_{min}, and $q^{(B)}_{\mathrm{th}}$ change also over time.

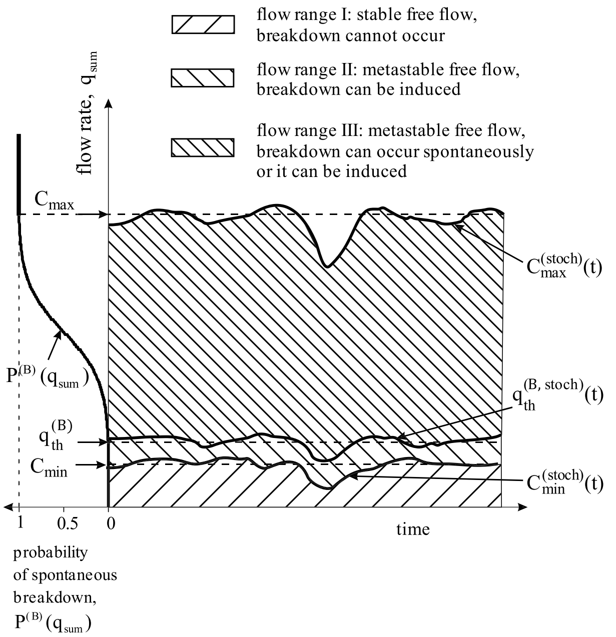

Fig. 13. Qualitative explanation of the infinite number of capacities of free flow at a highway bottleneck in three-phase traffic theory. Probability function for traffic breakdown $P^{(B)}(q_{sum})$ (*figure left*) is the *same* as that shown in Figs. 4(a), 8, and 12(b). Flow rate regions I, II, and III mentioned in labeling are the same as those shown in Fig. 4(a) and explained in Sect. 6 of the main text. Taken from [68]

Moreover, in real traffic flow, the traffic parameters are *stochastic* time functions.

Therefore, in real traffic flow we should consider some stochastic maximum capacity $C_{max}^{(stoch)}(t)$, stochastic minimum capacity $C_{min}^{(stoch)}(t)$, and a stochastic threshold flow rate $q_{th}^{(B,stoch)}(t)$ whose time dependence is determined by stochastic characteristics of traffic parameters. Qualitative hypothetical fragment of these time-functions within a time interval is shown in Fig. 13.

Stochastic functions $C_{max}^{(stoch)}(t)$, $C_{min}^{(stoch)}(t)$, and $q_{th}^{(B,stoch)}(t)$ shown in Fig. 13 are qualitative *hypothetical* functions that cannot be measured in empirical observations. Only their mean values (respectively, C_{max}, C_{min}, and $q_{th}^{(B)}$) can be found in empirical studies of measured traffic data. In particular, the mean values C_{max} and $q_{th}^{(B)}$ can be found from an empirical study of the flow rate function of the breakdown probability $P^{(B)}(q_{sum})$ (Fig. 4(a)).

It must be noted that in empirical observations the mean value of the minimum capacity C_{min} can be found from a study of a *finite number* of different days at which induced traffic breakdowns have been observed at a given bottleneck. The value C_{min} is related to these empirical days of observations only. In other words, it can occur that at another day, which is not within the days used for the calculation of C_{min}, traffic breakdown at this bottleneck can be induced at a smaller flow rate than the minimum capacity C_{min} found before. A similar comment is related to the physical meaning of the mean value of $q_{th}^{(B)}$. To explain this, we should note that with a finite number of measurements it is not possible to find some "exact value" of the minimum flow rate at which traffic breakdown can occur.

In other words, strictly speaking, mean values C_{min}, C_{max}, and $q_{th}^{(B)}$ are valid only for the days of the observing of traffic breakdown that have been used for the calculations of these mean values.

From Fig. 13 we can see that in three-phase traffic theory traffic breakdown *cannot* occur spontaneously at "any flow rate". Indeed,

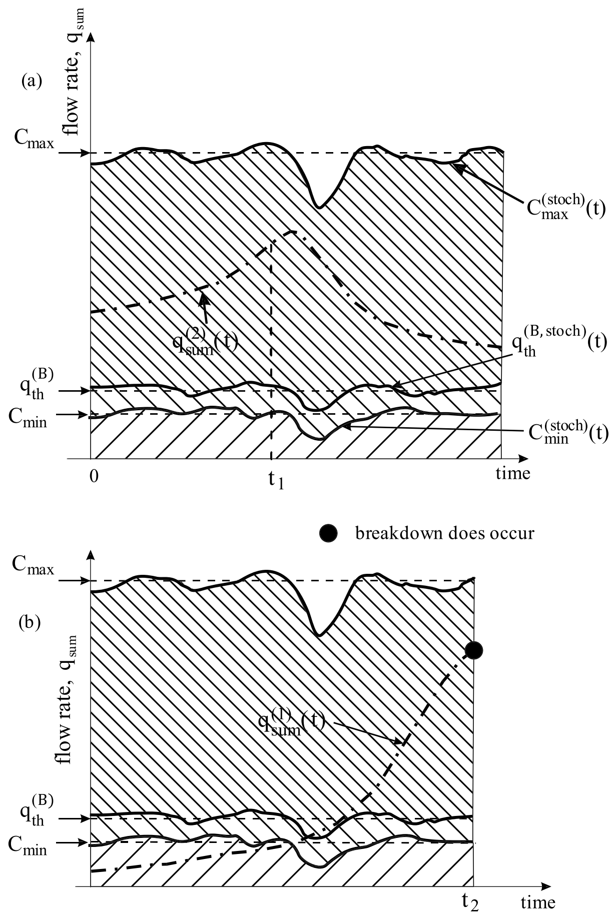

Fig. 14. Qualitative explanation of traffic breakdown with the use of the infinite number of capacities of free flow at a highway bottleneck of three-phase traffic theory [68]. Hypothetical time-functions $C_{max}^{(stoch)}(t)$, $C_{min}^{(stoch)}(t)$, and $q_{th}^{(B,stoch)}(t)$ are taken from Fig. 13. Hypothetical time functions of the flow rates $q_{sum}(t) = q_{sum}^{(2)}(t)$ in (a) and $q_{sum}(t) = q_{sum}^{(1)}(t)$ in (b) as well as time instant t_1 in (a) are, respectively, the same as those in Fig. 9

at any time when the flow rate in free flow is smaller than the minimum capacity $C_{min}^{(stoch)}(t)$, no traffic breakdown can occur at the bottleneck. When the flow rate $q_{sum}(t)$ satisfies conditions (4), specifically, $C_{min}^{(stoch)}(t) \le q_{sum}(t) < q_{th}^{(B,stoch)}(t)$, traffic breakdown can be induced only. Only under conditions $q_{th}^{(B,stoch)}(t) \le q_{sum}(t) < C_{max}^{(stoch)}(t)$ traffic breakdown can occur spontaneously with some probability $0 < P^{(B)}(q_{sum}) < 1$ during a given observation time.

Thus, we can see in Fig. 13 that in accordance with the highway capacity definition made in three-phase traffic theory, under conditions $C_{min}^{(stoch)}(t) \le q_{sum}(t) < C_{max}^{(stoch)}(t)$ *at any time instant* there is the infinite number of highway capacities at which traffic breakdown can occur with some probability or can be induced at the bottleneck.

Appendix D: Classical understanding of stochastic highway capacity versus infinite number of stochastic highway capacities of three-phase traffic theory

The objective of this appendix is to make a critical analysis of the classical definition of stochastic highway capacity that is generally accepted by most of the traffic researches (see references in the book [20]). We follow the associated critical analysis of the understanding of stochastic highway capacity made recently in [68].

The classical understanding of the nature of stochastic highway capacity (Appendix A) [20, 27–29] is based on the assumption that the empirical probability of traffic breakdown is determined by the capacity distribution function, i.e., that condition (9) is valid.

In contrast, the assumption of three-phase traffic theory about the metastability of traffic breakdown with respect to traffic breakdown (condition (10) of Appendix C) is based on the empirical evidence that traffic breakdown can be induced at a bottleneck (Appendix B).

As mentioned in Appendix A, the observation of empirical induced breakdowns proves that condition (9) of Brilon's stochastic capacity [20, 27–29] cannot be valid for real traffic. However, the following question arises:

- What are the consequences of this controversial understanding of the nature of traffic breakdown?

With the use of Fig. 13, we can qualitatively illustrate in Fig. 14 the basic difference between the classical understanding of the nature of stochastic highway capacity (Appendix A) and the understanding of the infinite number of stochastic highway capacities made in three-phase traffic theory (Appendix C).

In the classical understanding of stochastic capacity (Appendix A), for the hypothetical time dependence of the flow rate $q_{\mathrm{sum}}(t) = q^{(2)}_{\mathrm{sum}}(t)$ shown in Fig. 9(b), traffic breakdown has occurred at time instant t_1 at which $q^{(2)}_{\mathrm{sum}}(t_1) = C(t_1)$, i.e., when the flow rate is equal to the capacity value. In contrast, in three-phase traffic theory for the same time dependence of the flow rate $q^{(2)}_{\mathrm{sum}}(t)$, for which conditions $C^{(\mathrm{stoch})}_{\min}(t) \le q^{(2)}_{\mathrm{sum}}(t) < C^{(\mathrm{stoch})}_{\max}(t)$ are satisfied, no breakdown should be necessarily occur both at time instant t_1 and for a later time interval (Fig. 14(a)).

In the classical understanding of stochastic capacity (Appendix A), for the hypothetical time dependence of the flow rate $q_{\mathrm{sum}}(t) = q^{(1)}_{\mathrm{sum}}(t)$ shown in Fig. 9(a), traffic breakdown could not occur because for all time instants $q^{(1)}_{\mathrm{sum}}(t) < C(t)$. In contrast, in three-phase traffic theory for the same time dependence of the flow rate $q^{(1)}_{\mathrm{sum}}(t)$ traffic breakdown can occur spontaneously as this is shown for time instant t_2 in Fig. 14(b).

Because the classical understanding of stochastic highway capacity (8), (9) contradicts the empirical nucleation nature of real traffic breakdown, the understanding of stochastic highway capacity made in traditional traffic research community [20, 27–29] cannot be used for reliable highway design and highway operations.

Author

Boris S. Kerner
ist Experte für intelligente Transportsysteme, Verkehrsingenieurwesen und die Physik von Transport und Verkehr. Er wurde 1947 in Moskau, Sowjetunion, geboren und schloss 1972 sein Studium an der Technischen Universität Moskau MIREA ab. Er hat 1979 an der Russischen Akademie der Wissenschaft in Physik und Mathematik promoviert und sich 1986 habilitiert. Zwischen 1972 und 1992 beschäftigte er sich mit der Physik von Halbleitern sowie Plasma- und Festkörperphysik. Zwischen 1992 und Januar 2013 arbeitete Boris Kerner bei der Daimler AG in Stuttgart (seit 2000 als Leiter des wissenschaftlichen Fachgebietes *Verkehr*), wo sein Hauptinteresse dem Verständnis der real gemessenen Daten des Straßenverkehrs und der Entwicklung der intelligenten Transportsysteme gegolten hat. Die Erkenntnisse über die empirisch gemessenen Eigenschaften von räumlich-zeitlichen Mustern des Verkehrs bilden die Grundlage für die Drei-Phasen-Verkehrstheorie. Seit seiner Pensionierung im Februar 2013 arbeitet Boris Kerner als apl. Prof. an der Universität Duisburg-Essen im Bereich Physik von Transport und Verkehr.