Microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks

Boris S. Kerner[1] and Sergey L. Klenov[2]
[1]*DaimlerChrysler AG, RIC/TS, HPC: T729, 70546 Stuttgart, Germany*
[2]*Moscow Institute of Physics and Technology, Department of Physics, 141700 Dolgoprudny, Moscow Region, Russia*
(Received 30 January 2003; published 26 September 2003)

A microscopic theory of spatial-temporal congested traffic patterns at highway bottlenecks due to on-ramps, merge bottlenecks (a reduction of highway lanes), and off-ramps is presented. The basic postulate of three-phase traffic theory is used, which claims that homogeneous (in space and time) model solutions (steady states) of synchronized flow cover a two dimensional region in the flow-density plane [B. S. Kerner, Phys. Rev. Lett. **81**, 3797 (1998); Trans. Res. Rec. **1678**, 160 (1999)]. Phase transitions leading to diverse congested patterns, pattern evolution, and pattern nonlinear features have been found. Diagrams of congested patterns, i.e., regions of the pattern emergence dependent on traffic demand, have been derived. Diverse effects of metastability with respect to the pattern formation have been found. The microscopic theory allows us to explain the main *empirical* pattern features at on-ramps and off-ramps which have recently been found [B. S. Kerner, Phys. Rev. E **65**, 046138 (2002)]. (i) Rather than moving jams, synchronized flow first occurs at bottlenecks if the flow rate is slowly increasing. Wide moving jams can spontaneously occur only in synchronized flow. (ii) General patterns (GP) and synchronized flow patterns (SP) can spontaneously emerge at the bottlenecks. There can be the widening SP (WSP), the moving SP (MSP), and the localized SP. (iii) At on-ramps cases of ''weak'' and ''strong'' congestion should be distinguished. In contrast to weak congestion, under strong congestion the flow rate in synchronized flow in GP reaches a limit flow rate, the frequency of the moving jam emergence reaches a maximum, i.e., the GP characteristics under strong congestion do not depend on traffic demand. (iv) At the off-ramp GP with weak congestion occur. (v) A study of the pattern formation on a highway with two bottlenecks shows that diverse expanded patterns can occur, which cover both bottlenecks. SP first emerged at the downstream bottleneck can be caught at the upstream bottleneck (the catch effect). MSP, WSP, or wide moving jams first emerged at the downstream bottleneck induce diverse patterns at the upstream bottleneck. The onset of congestion at the upstream bottleneck can lead to an intensification of congestion at the downstream bottleneck. This causes a change in the pattern type and/or the pattern features.

DOI: 10.1103/PhysRevE.68.036130 PACS number(s): 89.40.−a, 47.54.+r, 64.60.Cn, 05.65.+b

I. INTRODUCTION

Traffic flow can be either free or congested (e.g., Refs. [1–28]). Congested patterns usually occur at highway bottlenecks (e.g., Ref. [19]) which act like defects for the pattern emergence in physical systems. In congested traffic two different traffic phases should be distinguished [22,25]: ''synchronized flow'' and ''wide moving jam.'' Thus, there are three traffic phases: (1) free flow, (2) synchronized flow, and (3) wide moving jam. *A moving jam* is restricted by two upstream moving fronts where the vehicle speed changes sharply. Inside the moving jam the speed is low and the vehicle density is high [20]. Empirical criteria for the traffic phases in congested traffic are the following [22,23,26]. The downstream front of a *wide* moving jam propagates on a highway *keeping* the mean velocity of the front, v_g. In contrast, the downstream front of synchronized flow (where vehicles accelerate escaping from synchronized flow to free flow) is usually fixed at the bottleneck (see Ref. [17]).

Theories of congested patterns remain controversial (see, e.g., Refs. [2–17,28]). Almost all traffic flow theories and models for freeway traffic up to now [14–16] are constructed such that in the unperturbed, noiseless limit the whole multitude of the hypothetical *steady* state model solutions where all vehicles move at the same distances to one another and with the same time-independent vehicle speed (these model solutions are also called ''homogeneous'' or ''equilibrium'' states; we will use the term ''steady'' states) is related to a curve(s) in the flow density plane. This curve(s) is called the *theoretical* fundamental diagram. There is a common feature of all models within this ''fundamental diagram approach'' which are claimed to show the spontaneous moving jam formation: If the flow rate upstream of the on-ramp is high enough and the flow rate to the on-ramp gradually increases beginning from zero, then first one moving jam and further a sequence of moving jams spontaneously occur in an initial free flow at the on-ramp [7–9,15,16].

By contrast, in empirical observations moving jams do *not* emerge spontaneously in free flow. Instead, first the transition from free flow to synchronized flow occurs (the $F \rightarrow S$ transition) at the on-ramp [17,25]. If the speed in synchronized flow is high enough then wide moving jams do *not* necessarily emerge in this synchronized flow [27]. In this case, synchronized flow patterns (SPs) occur at the bottleneck [17]. SP is a congested pattern at the bottleneck where no wide moving jams spontaneously emerge, i.e., SP consists only of synchronized flow. Wide moving jams can spontaneously emerge only in synchronized flow (the $S \rightarrow J$ transition) when the density in synchronized flow is high enough (the pinch effect [25]). Thus, wide moving jams emerge due to the sequence of two phase transitions: First the $F \rightarrow S$ transition and later, and on other location, the $S \rightarrow J$ transition (the $F \rightarrow S \rightarrow J$ transitions). As a result, the general pattern (GP) occurs at a bottleneck [17]: GP is the congested

pattern at the bottleneck where synchronized flow occurs upstream of the bottleneck and wide moving jams spontaneously emerge in that synchronized flow.

To explain these empirical results, Kerner introduced three-phase traffic theory. The fundamental *hypothesis* of this theory is that steady states of synchronized flow cover *a two-dimensional (2D) region* in the flow-density plane, i.e., there is *no* fundamental diagram for steady states of synchronized flow in this theory [23–25]. As a proof of three-phase traffic theory its results may be considered (see also Refs. [12,28]) which allow us to overcome the mentioned problems of fundamental diagram approach for the description of congested traffic.

The above critical consideration of the fundamental diagram approach (see also Refs. [27]) does not concern important ideas introduced in the related theories for the modeling of vehicle safety conditions, fluctuations, vehicle acceleration and deceleration, and different vehicle time delays (e.g., Refs. [5,6,10,14–16,18,29]). These effects are very important elements of a microscopic three-phase traffic theory in Refs. [12,28] and in the present paper.

Recently features of empirical spatial-temporal congested patterns at an isolated bottleneck (the bottleneck which is far enough from other bottlenecks) and of the patterns covering several bottlenecks have been found out [17]. SPs and GPs at an isolated on-ramp have been studied in a microscopic three-phase traffic theory [12]. However, there is a diverse variety of nonlinear empirical pattern features [17] which have not been found in theories. In this paper, a microscopic theory of these diverse pattern features is presented.

The paper is organized as follows. A microscopic traffic flow model will be formulated in Sec. II. Features of the breakdown phenomenon (the $F \rightarrow S$ transition) and of the moving jam emergence will be found in Secs. III and IV. In Secs. V, VII, and VIII, strong congestion and weak congestion conditions at different bottlenecks will be studied. Metastability and hysteresis effects will be examined in Secs. VI and VII. Induced pattern formation, the catch effect, expanded patterns, pattern interaction, and an intensification of the downstream congestion due to the upstream congestion which occur when two different bottlenecks exist on a highway will be considered in Secs. IX, X, and XI. A comparison of empirical results with the theory is made in Secs. V E and XII.

II. A MICROSCOPIC TRAFFIC FLOW MODEL FOR SPATIAL-TEMPORAL CONGESTED PATTERNS

A. General rules of vehicle motion

In a one-lane traffic flow model the general rules of vehicle motion first introduced in Ref. [12] are used:

$$v_{n+1} = \max(0, \min(v_{\text{free}}, v_{c,n}, v_{s,n})), \tag{1}$$

$$x_{n+1} = x_n + v_{n+1}\tau, \tag{2}$$

$$v_{c,n} = \begin{cases} v_n + \Delta_n & \text{at } x_{\ell,n} - x_n \leqslant D_n \\ v_n + a_n\tau & \text{at } x_{\ell,n} - x_n > D_n, \end{cases} \tag{3}$$

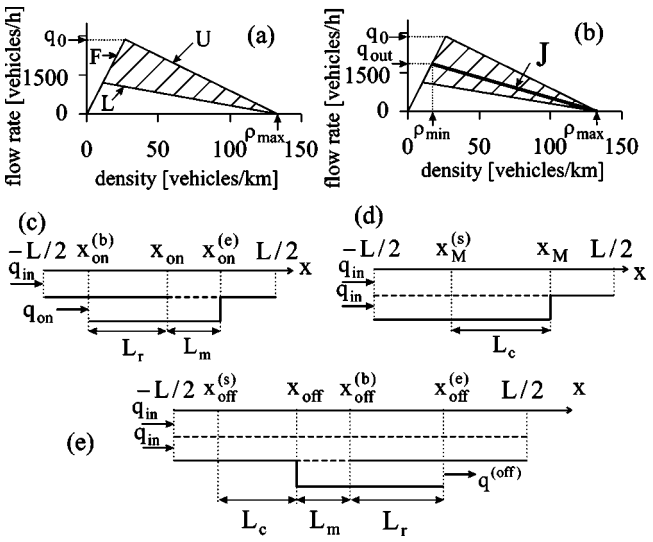

FIG. 1. Steady states of the model in the flow-density plane (a), the line J (b) and the models of the on-ramp (c), the merge bottleneck (d), and the off-ramp (e). In (a), (b) the two-dimensional region in the flow-density plane (dashed region) is limited by three boundaries U, L, and F which have been explained in Refs. [12,28]. At chosen model parameters (Sec. II C) when free flow is formed in the outflow from a wide moving jam, the outflow rate is $q_{\text{out}} = 1810$ vehicles/h; the downstream jam front velocity which determines the slope of the line J in (b) is $v_g = -15.5$ km/h.

where v_n and x_n are the speed and the space coordinate of the vehicle; the index n corresponds to the discrete time $t = n\tau$, $n = 0,1,2,\ldots$; τ is the time step; v_{free} is the maximum speed in free flow which is considered as a constant value; $v_{s,n}$ is the save speed; $v_{c,n}$ is a desirable speed, Δ_n is given by the formula

$$\Delta_n = \max(-b_n\tau, \min(a_n\tau, v_{\ell,n} - v_n)), \tag{4}$$

$a_n \geqslant 0$ is acceleration, $b_n \geqslant 0$ is deceleration; the lower index ℓ marks variables related to the vehicle in front of that at x_n, the "leading vehicle"; all vehicles have the same length d; D_n is "the synchronization distance": At $x_{\ell,n} - x_n \leqslant D_n$ the vehicle tends to adjust its speed to the speed of the leading vehicle, i.e., the vehicle decelerates if $v_n > v_{\ell,n}$, and accelerates if $v_n < v_{\ell,n}$ [12,28]. The synchronization distance, as has been shown in Refs. [12,28], is related to the fundamental hypothesis of three-phase traffic theory [23–25]. Hypothetical steady states of synchronized flow cover a 2D region in the flow-density plane [Figs. 1(a,b)].

In the general rules (1)–(4) D_n is chosen as

$$D_n = d + \max(0, kv_n\tau + \beta a^{-1} v_n (v_n - v_{\ell,n})), \tag{5}$$

where $k > 1$, β, and a are constants. If $v_n = v_{\ell,n}$ the synchronization distance D_n is $d + kv_n\tau$. This corresponds to a fixed time gap $k\tau$. If $v_n > v_{\ell,n}$, the distance D_n increases and vice versa.

The safe speed $v_{s,n}$ in Eq. (1) is taken for a homogeneous one-lane road from Ref. [5] and it has been modified for a road with bottlenecks [Figs. 1(c)–1(e)] to take into account the anticipation effect [10,11]. We use

$$v_{s,n}=\min(v_n^{(safe)},g_n/\tau+v_\ell^{(a)}), \tag{6}$$

where $v_n^{(safe)}=v^{(safe)}(g_n,v_{\ell,n})$ is given by formulas (3)–(5) in Ref. [5] with the deceleration $b=1$ m/s^2 and the space gap $g_n=x_{\ell,n}-x_n-d$; $v_\ell^{(a)}$ is an anticipation speed chosen as $v_\ell^{(a)}=\max(0,\min(v_{\ell,n}^{(safe)}-a\tau,v_{\ell,n}-a\tau,g_{\ell,n}/\tau))$.

B. Random acceleration and deceleration

As well as in other models [6,10,14,15], we use variables which are stochastic functions. *At the first step*, the preliminary speed of each vehicle is found as $\tilde{v}_{n+1}=v_{n+1}$ where v_{n+1} is calculated based on Eqs. (1)–(4) combined with Eqs. (5) and (6). *At the second step*, a noise component ξ_n is added to the calculated speed $\tilde{v}_{n+1}$ and then the final value of the speed v_{n+1} at the time $n+1$ is found from the condition introduced in Ref. [28],

$$v_{n+1}=\max(0,\min(v_{\text{free}},\tilde{v}_{n+1}+\xi_n,v_n+a\tau,v_{s,n})). \tag{7}$$

1. "Continuous" noise model

First we considered a model in which $a_n=b_n=a$ in Eq. (4), and ξ_n in Eq. (7) is given as

$$\xi_n=a\tau\,\text{rand}(-\epsilon_b,\epsilon_a), \tag{8}$$

where rand(y,z) denotes a random value uniformly distributed between y and z. In (8) the approach to noise modeling from Refs. [5,30] is used. In comparison with Refs. [5,30], Eq. (8) also describes a random acceleration; besides, values $0\leqslant\epsilon_a,\epsilon_b\leqslant 1$ in Eq. (8) can be functions of the speed [31].

Pattern features can be found with model (1)–(8). However, some unrealistic pattern features appear (e.g., like as in Ref. [28] the velocity of the upstream front of synchronized flow can be a nonrealistic one). To have simulated pattern parameters close to the empirical ones [17], we have used the following "motion state model" in all simulations presented in the paper.

2. Motion state model

In the model, a random deceleration and acceleration are applied depending on whether the vehicle decelerates or accelerates or else maintains its speed:

$$\xi_n=\begin{cases}-\xi_b & \text{if } S_{n+1}=-1\\ \xi_a & \text{if } S_{n+1}=1\\ 0 & \text{if } S_{n+1}=0,\end{cases} \tag{9}$$

where ξ_b and ξ_a are random sources for deceleration and acceleration, respectively; S in Eq. (9) denotes the state of motion ($S_{n+1}=-1$ is related to a deceleration, $S_{n+1}=1$ to an acceleration, and $S_{n+1}=0$ to the motion at nearly constant speed)

$$S_{n+1}=\begin{cases}-1 & \text{if } \tilde{v}_{n+1}<v_n-\delta\\ 1 & \text{if } \tilde{v}_{n+1}>v_n+\delta\\ 0 & \text{otherwise,}\end{cases} \tag{10}$$

where δ is a constant ($\delta\ll a\tau$).

The random components in Eq. (9) are given as "impulsive" ones: $\xi_b=a\tau\theta(p_b-r)$ and $\xi_a=a\tau\theta(p_a-r)$ where p_b and p_a are, respectively probabilities of random deceleration and acceleration, $r=\text{rand}(0,1)$, $\theta(z)=0$ at $z<0$ and $\theta(z)=1$ at $z\geqslant 0$.

To simulate a delay time either in the vehicle acceleration or in the vehicle deceleration [14,15], a_n and b_n in Eq. (4) are taken as the following stochastic functions:

$$a_n=a\,\theta(P_0-r_1),\quad b_n=a\,\theta(P_1-r_1), \tag{11}$$

$$P_0=\begin{cases}p_0 & \text{if } S_n\neq 1\\ 1 & \text{if } S_n=1,\end{cases}\quad P_1=\begin{cases}p_1 & \text{if } S_n\neq -1\\ p_2 & \text{if } S_n=-1,\end{cases} \tag{12}$$

where $r_1=\text{rand}(0,1)$; $1-P_0$ and $1-P_1$ are, respectively, the probabilities of random delay in the vehicle acceleration and deceleration at time step $n+1$.

The first formulas in Eqs. (11) and (12) together with Eqs. (3) and (4) simulate the delay in acceleration that takes place if the vehicle does not accelerate ($S_n\neq 1$) at the time step n. The mean delay time after which the vehicle starts to accelerate is $\tau_{del}^{(a)}=\tau/p_0(v_n)$. The delay time is supposed to be maximal for vehicles in a standstill ($v_n=0$). This is achieved by the choice of probability p_0 as an increasing function $p_0(v_n)$ of the speed (Sec. II C). At $v_n=0$ the first formulas in Eqs. (11) and (12) together with Eqs. (3) and (4) simulate the slow-to-start rules [6,29]. Vehicles escape at the downstream front of a wide moving jam with the mean delay time $\tau_{del}^{(a)}(0)=\tau/p_0(0)$ [11]. This gives the velocity of the jam downstream front $v_g=-p_0(0)d/\tau$ corresponding to the formula $v_g=-1/[\rho_{max}\tau_{del}^{(a)}(0)]$ [25], $\rho_{max}=d^{-1}$ is the density inside the jam.

The second formulas in Eqs. (11) and (12) together with Eq. (4) simulate the delay in deceleration which occurs if $v_n>v_{\ell,n}$ and $x_{\ell,n}-x_n\leqslant D_n$. The vehicle which is not decelerating at time step n ($S_n\neq -1$) starts to decelerate at time step $n+1$ with the probability p_1. The mean delay time of the deceleration start is $\tau_{del}^{(d)}=\tau/p_1$. However, if the vehicle has already been decelerating at time step $n(S_n=-1)$, it continues to decelerate at time step $n+1$ with probability p_2 and interrupts decelerating with probability $1-p_2$. The latter case corresponds to the situation that a driver closes up to the leading vehicle to minimize the space gap. This effect is more essential for vehicles moving at low speed in the dense flow that allows us to simulate the pinch effect, i.e., the self-compression and narrow moving jam emergence in synchronized flow [25]. Therefore, the probability p_2 is taken as an increasing function $p_2(v_n)$ of the speed (Sec. II C).

C. Boundary conditions, models of bottlenecks, and simulation parameters

In model (1)–(7),(9)–(12), in the incoming boundary flow with the flow rate q_{in}, a new vehicle is generated at the start of the road at time interval $\tau_{in}=1/q_{in}$ after generation of the last vehicle, provided the distance to the last vehicle exceeds the safe one $v_n\tau+d$. The speed of each new generated ve-

hicle is set to $v_n = v_{\ell,n}$. After a vehicle has reached the end of the road it is removed; before this the most downstream vehicle maintains its speed.

The bottleneck due to an on-ramp, the merge bottleneck, where two lanes are reduced to one lane, and due to an off-ramp [Figs. 1(c)–1(e)] are considered. The on-ramp consists of two parts [Fig. 1(c)]: (i) the merging region of length L_m where a vehicle may squeeze to the main road and (ii) the part of the on-ramp upstream of the merging region (the on-ramp lane with length L_r) where vehicles move according to model (1)–(7),(9)–(12) with the maximal speed $v_{\text{free, on}} = 80$ km/h. At the start of the on-ramp lane ($x = x_{\text{on}}^{(b)}$) the flow rate to the on-ramp q_{on} is given in the same manner as q_{in}.

In the merge bottleneck [Fig. 1(d)], within the merging region of length L_c upstream of the merge point $x = x_M$ the vehicles have to change from the right lane to the left lane.

The off-ramp consists of two parts [Fig. 1(e)]: (i) The merging region of length L_m where a vehicle may squeeze from the main road to the off-ramp lane and (ii) the off-ramp lane of length L_r downstream of the merging region where vehicles move according to model (1)–(7),(9)–(12) with the maximal speed $v_{\text{free,off}} = 90$ km/h. Within a second merging region of the length $L_m + L_c$ [$x_{\text{off}}^{(s)} \leqslant x \leqslant x_{\text{off}}^{(b)}$ in Fig. 1(e)] vehicles going to the off-ramp have to change from the left lane to the right lane of the main road. The flow rate of vehicles which want to move from the road to the off-ramp is given as a percent η of the flow rate q_{in}.

The lane changing rules from Ref. [12] are used beyond all bottleneck merging regions. The rules for vehicle squeezing within the merging regions are assumed to be the same for all three types of bottlenecks. Rule (A): The security lane changing rules (19) from Ref. [12] are fulfilled. After the squeezing the coordinate of the vehicle on the new lane remains the same. Rule (B): The gap $x_n^+ - x_n^- - d$ between two neighboring vehicles on the other lane exceeds some value $g_{\text{on}}^{(\min)} = \lambda v_n^+ + d$, where λ is a parameter [upper indices "+" "(−)" belong to the vehicle ahead (behind)]. In addition, the vehicle passes the middle point $x^{(m)} = (x_n^+ + x_n^-)/2$ for the time step n. In this case the coordinate of the squeezing vehicle is set to $x = x^{(m)}$. It is also assumed that before squeezing the vehicle tries to adjust its speed to the value $v_n^+ + \Delta v_r^{(2)}$. After squeezing the vehicle changes its speed by a finite value: $v_n \to \min(v_n^+, v_n + \Delta v_r^{(1)})$, where $\Delta v_r^{(1)}$, $\Delta v_r^{(2)}$ are constants. The new value of v_n is used in rule (A).

In simulations, the length of the main road is $L = 40$ km, the point $x = 0$ is at the distance $L/2 = 20$ km from the end of the road so the road starts at $x = -L/2 = -20$ km. The model parameters are $\tau = 1$ s, $v_{\text{free}} = 30$ m/s (108 km/h), $d = 7.5$ m, $a = 0.5$ m/s^2, $k = 3$, $\beta = 1$, $p_1 = 0.3$, $p_a = 0.17$, $p_b = 0.1$, $\delta = 0.01$, $p_0(v) = 0.575 + 0.125 \min(1, v/10)$, $p_2(v) = 0.48 + 0.32\theta(v - 15)$. Lane changing parameters are the same as in Ref. [12] except $\gamma^- = 1$. The parameters of the merging region are $\lambda = 0.75$ ($\lambda = 0.6$ for the merging region on the off-ramp lane), $\Delta v_r^{(1)} = 10$ m/s, $\Delta v_r^{(2)} = 5$ m/s at the on-ramp and $\Delta v_r^{(2)} = -2$ m/s at the off-ramp and at the merge bottleneck. $x_{\text{on}} = x_M = x_{\text{off}} = 16$ km. $L_m = 300$ m, $L_r = 2$ km for the on-ramp; $L_m = 0.6$ km, $L_r = 2$ km, $L_c = 1.6$ km for the off-ramp; $L_c = 0.6$ km for the merge bottleneck. When other values of model parameters are used for some simulations, they are given in the related figure captions.

III. THE BREAKDOWN PHENOMENON (THE $F \to S$ TRANSITION)

A. Homogeneous road: The moving synchronized flow pattern

In model (1)–(7),(9)–(12), as well as in Ref. [12], rather than moving jams the $F \to S$ transition (the breakdown phenomenon) can occur on a homogeneous (without bottlenecks) one-lane road. We have found the following features.

(i) Within the range

$$q_{\text{th}} \leqslant q_{\text{in}} \leqslant q_{\max} \quad (\rho_{\text{th}} \leqslant \rho_{\text{in}} \leqslant \rho_{\max}^{(\text{free})}), \tag{13}$$

where $\rho_{\text{in}} = q_{\text{in}}/v_{\text{free}}$, the nucleation of synchronized flow in an initial free flow can occur, i.e., the free flow is metastable with respect to the first order $F \to S$ transition. A Z-shaped speed-flow characteristic [Fig. 2(a)] is related to these metastability and nucleation effects. The Z characteristic consists of the branch of free flow v_{free}, of a 2D region of synchronized flow states v_{syn} and of the critical branch $v_{\text{cr}}^{(\text{FS})}$ which gives the speed inside the critical perturbations, i.e., it determines the critical amplitude of local perturbations in free flow $\Delta v_{\text{cr}}^{(\text{FS})} = v_{\text{free}} - v_{\text{cr}}^{(\text{FS})}$. When the amplitude of an external local perturbation exceeds $\Delta v_{\text{cr}}^{(\text{FS})}$ the $F \to S$ transition [down arrow in Fig. 2(a)] occurs and a moving synchronized flow pattern (MSP) emerges [Figs. 2(b)–2(d)] whose width continuously increases over time. Otherwise [up arrow in Fig. 2(a)] the initial perturbation decays [Fig. 2(e)]. In the speed-density plane, there is a gap in the density between states v_{free} and states v_{syn} inside MSP [Fig. 2(f)]. At $q = q_{\max}$ the $F \to S$ transition [dotted down arrow in Fig. 2(a)] occurs already at a very small amplitude of perturbations.

(ii) The $F \to S$ transition and the Z-shaped speed-flow characteristic may be explained by a competition between a tendency to free flow due to an "overacceleration" which is simulated by the acceleration noise in (9) and a tendency to synchronized flow due to the adaptation of the vehicle speed to the speed of the leading vehicle. This adaptation is simulated with Eqs. (3) and (4) due to the synchronization distance D_n and the deceleration noise in Ref. (9). The "overacceleration" is stronger at higher vehicle speed, exactly lower density (the flow rate is nearly the same in free and synchronized flows [Figs. 2(c,d)]. This causes the $S \to F$ transition and the MSP dissolving [Fig. 2(e)]. In contrast, the tendency of speed adaptation is stronger at lower speed, i.e., at the higher density. This causes the self-maintenance of synchronized flow in MSP.

(iii) At the threshold point $(\rho_{\text{th}}, q_{\text{th}})$ [Figs. 2(a,f)] the downstream front velocity of MSP v_{down} is equal to the velocity of the MSP upstream front v_{up}: $v_{\text{down}} = v_{\text{up}}$ [Fig. 2(h)] [32]. At $\rho_{\text{in}} > \rho_{\text{th}}$ ($q_{\text{in}} > q_{\text{th}}$) we have $v_{\text{down}} > v_{\text{up}}$ [Fig. 2(h)], i.e., the MSP width increases over time [Fig. 2(b)]. In contrast, at $\rho_{\text{in}} < \rho_{\text{th}}$ ($q_{\text{in}} < q_{\text{th}}$) we have $v_{\text{down}} < v_{\text{up}}$ [Fig. 2(h)], i.e., the MSP width decreases and the MSP dissolves [Fig. 2(g)]. To explain Fig. 2(h), let us use the formulas v_{down}

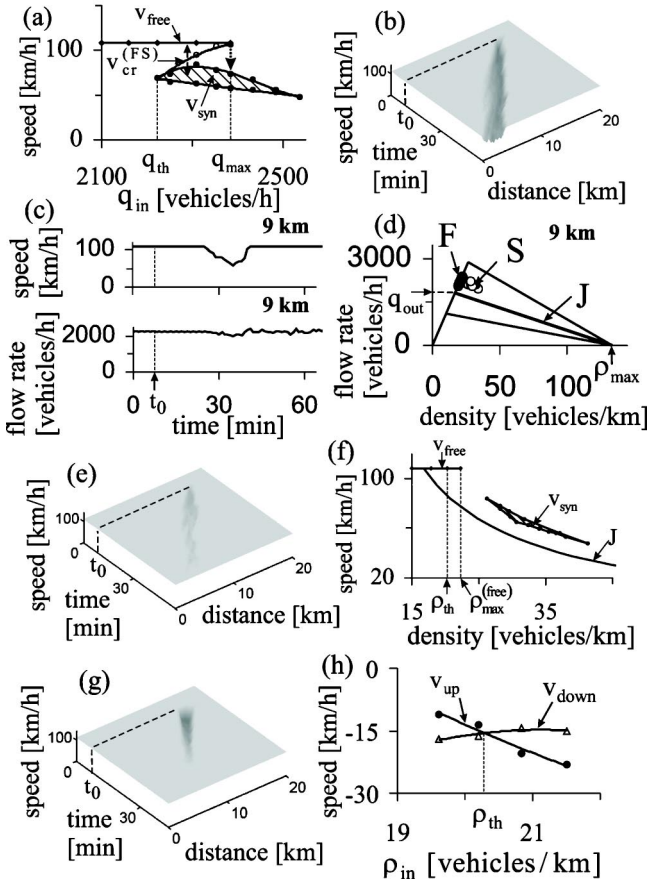

FIG. 2. The breakdown phenomenon on a homogeneous road. (a) A Z-shaped speed-flow characteristic. In (a) $q_{\rm in}$ is a ''control'' (external) parameter; the speed $v_{\rm free}$ is related to free flow; the critical branch $v_{\rm cr}^{\rm (FS)}$ gives the speed inside the critical local perturbation for the $F\rightarrow S$ transition as a function of $q_{\rm in}$; $v_{\rm syn}$ is a synchronized flow speed inside the 2D region of synchronized flow states (dashed region), $v_{\rm syn}$ is related to an averaging of speeds inside MSP over 10-min time intervals. (b) The speed in space and time during the emergence of MSP due to the growth of an external perturbation. The speed (top), the flow rate (bottom) (c) and the corresponding data in the flow-density plane (d) for MSP in (b). (e) The decay of an external perturbation whose amplitude is lower than $\Delta v_{\rm cr}^{\rm (FS)}$. (f) The speed-density characteristic where $v_{\rm free}$ and $v_{\rm syn}$ have the same meaning as in (a); the curve J is related to the line J in Fig. 1(b). (g) The speed in space and time during the decay of an external perturbation at $q_{\rm in}<q_{\rm th}$ ($\rho_{\rm in}<\rho_{\rm th}$). (h) Dependencies $v_{\rm down}$ and $v_{\rm up}$ on $\rho_{\rm in}=q_{\rm in}/v_{\rm free}$. In (c),(d) 1-min averaged data of virtual detectors are presented. In (d) black points and circles are related to free flow (F) and to synchronized flow (S) inside MSP (b), respectively. $q_{\rm in}=2250$ vehicles/h in (b)–(e) and $q_{\rm in}=2118$ vehicles/h in (g). In (a), (f) $\rho_{\rm th}\approx 20.4$ vehicles/h, $\rho_{\rm max}^{\rm (free)}\approx 22.5$ vehicles/km, the flow rates $q_{\rm th}$ and $q_{\rm max}$ related to $\rho_{\rm th}$ and $\rho_{\rm max}^{\rm (free)}$, respectively, are $q_{\rm th}\approx 2200$ vehicles/h and $q_{\rm max}\approx 2430$ vehicles/h. An external perturbation is applied at $t=t_0=8$ min and is given as a braking of one of the vehicles in an initial free flow up to some speed which is maintained then for 60 s in (a),(b),(e) and 150 s in (g).

$=v_{\rm syn}-[1/(\tau_{\rm del}^{\rm (a)}\rho_{\rm syn})]$, $v_{\rm up}=(q_{\rm syn}-q_{\rm in})/(\rho_{\rm syn}-\rho_{\rm in})$ where $q_{\rm syn}$, $\rho_{\rm syn}$ and $v_{\rm syn}$ are, respectively, the averaged flow rate, density, and speed inside MSP [Figs. 2(b,c)]. We see that

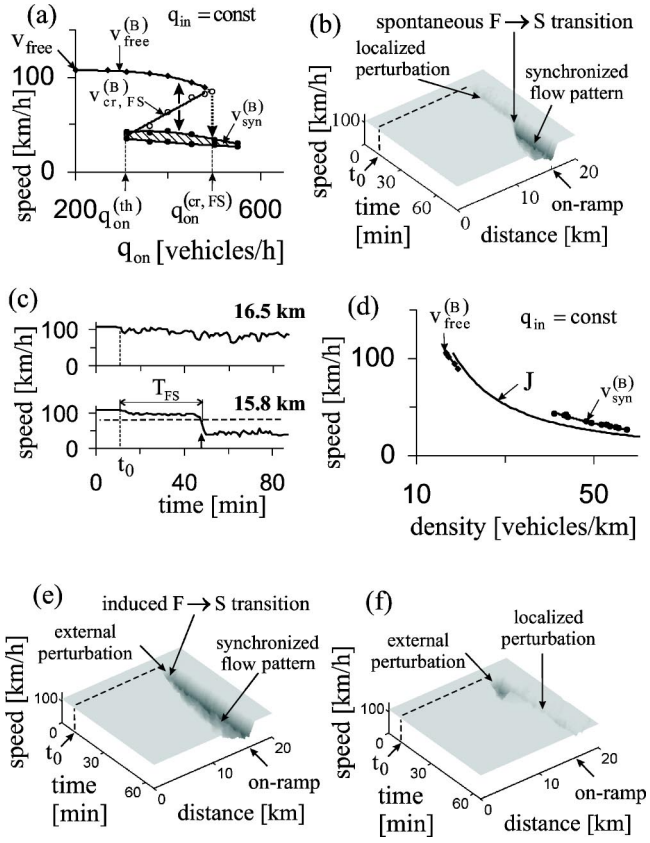

FIG. 3. The breakdown phenomenon (the $F\rightarrow S$ transition) at the on-ramp. The Z-shaped dependence of the speed on $q_{\rm on}$ (a ''control'' parameter) at a fixed $q_{\rm in}$ (a), the speed in space and time (b) and at virtual detectors (c) during the spontaneous $F\rightarrow S$ transition, and the speed-density characteristic (d). In (a) the speed $v_{\rm free}^{\rm (B)}$ is related to free flow on the main road in the on-ramp vicinity; the critical branch $v_{\rm cr,\,FS}^{\rm (B)}$ gives the speed inside the critical local perturbation for the $F\rightarrow S$ transition as a function of $q_{\rm on}$. In (a),(d) $v_{\rm syn}^{\rm (B)}$ is a synchronized flow speed inside the 2D region (dashed region; 10-min averaged data of a virtual detector on the main road at $x=16$ km). The curve J is related to the line J in Fig. 1(b). The speed in space and time during the induced SP emergence due to an external perturbation (e) and the decay of an external perturbation (f). In (c) the dashed horizontal line shows the speed level 85 km/h; the criterion for the $F\rightarrow S$ transition is the same as in Ref. [28]. In (a),(d) $q_{\rm in}=1756$ vehicles/km. In (b),(c),(e),(f) $q_{\rm in}=1800$ vehicles/h, $q_{\rm on}$ is 450 (b),(c) and 400 (e),(f) vehicles/h. In (a), to find $v_{\rm cr,\,FS}^{\rm (B)}(q_{\rm on})$ a perturbation as in Figs. 2(a, b) is applied on the main road at $x=16$ km for 120 s. In (b,c,e,f) the on-ramp inflow $q_{\rm on}$ is switched on at $t=t_0=8$ min.

$v_{\rm down}$ does not directly depend on $q_{\rm in}$ whereas $v_{\rm up}$ decreases when $q_{\rm in}$ increases.

B. The breakdown phenomenon at the on-ramp

In model (1)–(7),(9)–(12) rather than moving jams the $F\rightarrow S$ transition can spontaneously occur at an on-ramp [12,28]. We have found the following features:

(i) The average speed in free flow $v_{\rm free}^{\rm (B)}<v_{\rm free}$. $v_{\rm free}^{\rm (B)}$ decreases when $q_{\rm on}$ increases [Fig. 3(a)]. This is linked to the spontaneous occurrence of an internal localized perturbation

["localized perturbation" in Fig. 3(b)] on the main road in the on-ramp vicinity caused by the vehicle squeezing. The perturbation amplitude oscillates over time: The perturbation consists of a permanent ("deterministic") and a random component.

(ii) At a given q_{in} in an initial free flow within a range $q_{on}^{(th)} \leqslant q_{on} \leqslant q_{on}^{(cr,FS)}$ [Fig. 3(a)] due to the self-growth of the internal perturbation [item (i)] the *spontaneous* nucleation of synchronized flow can occur: The free flow is metastable with respect to the first order spontaneous $F \rightarrow S$ transition leading to the SP emergence [Fig. 3(b)]. The SP nucleation occurs after a delay time T_{FS} [Fig. 3(c)] [28]. The metastability and nucleation effects are linked to the competition between the "overacceleration" and the speed adaptation in synchronized flow [Sec. III A, item (ii)]. These effects lead also to a Z-shaped speed-flow characteristic [Fig. 3(a)]. This speed Z-shaped function of q_{on} at a given q_{in} [Fig. 3(a)] consists of the branch of free flow $v_{free}^{(B)}$, of a 2D region of synchronized flow $v_{syn}^{(B)}$ (dashed region), and of the critical branch $v_{cr,FS}^{(B)}$which determines the critical amplitude $\Delta v_{cr,FS}^{(B)} = v_{free}^{(B)} - v_{cr,FS}^{(B)}$ of local perturbations: If the perturbation amplitude exceeds $\Delta v_{cr,FS}^{(B)}$, the $F \rightarrow S$ transition occurs [down arrow in Fig. 3(a)]. Otherwise [up arrow in Fig. 3(a)] perturbations decay. In the speed-density plane, there is a gap in the density between free flow and synchronized flow states [Fig. 3(d)].

(iii) The delay time T_{FS} [Fig. 3(c)] is a random value: In different realizations, at the same q_{on} and q_{in} we found usually different T_{FS}. The mean delay time $T_{FS}^{(mean)}$ is higher, the lower q_{on} and/or q_{in} are. At higher q_{in} during the delay time T_{FS} a lot of perturbations [item (i)] can spontaneously occur, grow, and then decay in the on-ramp vicinity before the perturbation appears whose growth leads to the $F \rightarrow S$ transition. The critical perturbation amplitude $\Delta v_{cr,FS}^{(B)}$ and therefore $T_{FS}^{(mean)}$ decrease when q_{on} increases. In the vicinity of $q_{on} = q_{on}^{(cr,FS)}$ the $F \rightarrow S$ transition occurs already after $T_{FS} \approx 1-2$ min that is comparable with the duration of the $F \rightarrow S$ transition. Thus, the exact value $q_{on} = q_{on}^{(cr,FS)}$ cannot be found [dashed parts of $v_{free}^{(B)}(q_{on})$ and $v_{cr,FS}^{(B)}(q_{on})$]. At $q_{on} \geqslant q_{on}^{(cr,FS)}$ the "deterministic" spontaneous $F \rightarrow S$ transition is expected [dotted down-arrow in Fig. 3(a)].

(iv) External local perturbations can lead to the induced $F \rightarrow S$ transition and SP emergence [Fig. 3(e)] when the amplitude of the external perturbation is higher than $\Delta v_{cr,FS}^{(B)}$. Otherwise the perturbation decays and free flow $v_{free}^{(B)}$ is recovered near the on-ramp [Fig. 3(f)].

IV. MOVING JAM EMERGENCE: DOUBLE Z-SHAPED CHARACTERISTICS OF TRAFFIC FLOW

Synchronized flow states v_{syn} in Fig. 2(f) and $v_{syn}^{(B)}$ in Fig. 3(d) lie above the curve *J*.Thus, wide moving jams can *spontaneously* emerge in these states [12,25,28]. We found the following features.

(i) For a homogeneous road, there is a Z-shaped speed-flow characteristic [Fig. 4(a)]: The *Z* characteristic is related to the metastability of synchronized flow v_{syn} and to the *spontaneous* nucleation effect in this flow leading to the first order $S \rightarrow J$ transition. The *Z*-characteristic consists of the states v_{syn} (dashed 2D region), of the critical branch $v_{cr}^{(SJ)}$ which determines the critical amplitude of local perturbations $\Delta v_{cr}^{(SJ)} = v_{syn} - v_{cr}^{(SJ)}$ [33], and of the line $v_{min} = 0$ for the speed inside a wide moving jam. The nucleation of the $S \rightarrow J$ transition occurs in the range $q_{th} \leqslant q_{in} \leqslant q_{cr}$ if the perturbation amplitude in synchronized flow exceeds $\Delta v_{cr}^{(SJ)}$ (down arrow); otherwise, the perturbation decays (up arrow). There is a random delay time T_{SJ} for the $S \rightarrow J$ transition occurrence. In the vicinity of $q_{on} = q_{cr}$ fluctuations cause the $S \rightarrow J$ transition during $T_{SJ} \approx 10$ min which is comparable

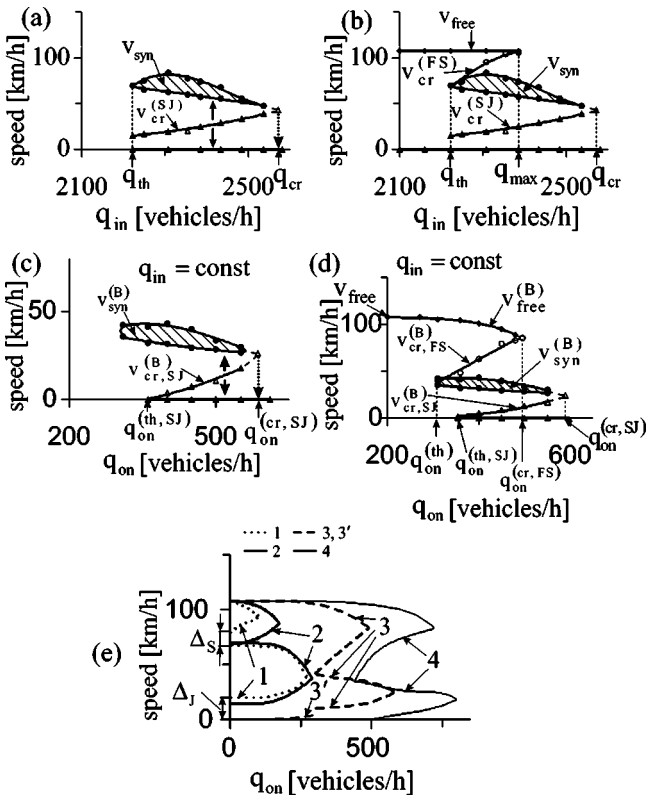

FIG. 4. *Z* characteristics for the $S \rightarrow J$ transition (a),(c) and double *Z* characteristics for the $F \rightarrow S \rightarrow J$ transitions (b),(d),(e). (a),(b) A homogeneous road. (c)–(e) A road with the on-ramp. In (a)–(d) the critical branches $v_{cr}^{(SJ)}$ (a),(b) and $v_{cr,SJ}^{(B)}$ (c),(d) give the speed inside the critical local perturbation for the $S \rightarrow J$ transition as a function of q_{in} (a),(b) and as a function of q_{on} (c),(d), respectively. The states v_{free}, $v_{cr}^{(FS)}$, and v_{syn} in (b) are taken from Fig. 2(a) and the states $v_{cr}^{(SJ)}$ are taken from (a). The states $v_{free}^{(B)}$, $v_{cr,FS}^{(B)}$, and $v_{syn}^{(B)}$ in (d) are taken from Fig. 3(a) and the states $v_{cr,SJ}^{(B)}$ are taken from (c); in (c),(d) q_{in} is a constant. The critical branch $v_{cr,SJ}^{(B)}(q_{on})$ (c),(d) is found by a braking of one of the vehicles in synchronized flow on the main road at $x = 15.7$ km up to the speed $v_{cr,SJ}^{(B)}$. In (e) double *Z* characteristics for different ranges of q_{in} are shown: curves 1 for $q_{in} > q_{out}$, q_{th}; curve 2 for $q_{in} = q_{th}$; curves 3,3′ for $q_{out} < q_{in} < q_{th}$; curve 4 for $q_{in} < q_{out}$. For a simplification, in (e) the branches of synchronized flow states $v_{syn}^{(B)}$ are related to an averaging of the initial infinity different speeds for a given q_{on} within the 2D region in (d). In (c)–(e) q_{in} is: in (c),(d) 1756 ($q_{in} < q_{out}$), in (e) 2265 (curves 1), 2200 (curve 2), 1850 (curves 3,3′), 1600 (curve 4) vehicles/km. In (e) the branch 3′ shows the critical branch $v_{cr,FJ}^{(B)}$ for the *induced* F→J transition [35].

with the duration of the $S \rightarrow J$ transition. The exact point $q_{in}=q_{cr}$ cannot be found (dashed parts of v_{syn} and $v_{cr}^{(SJ)}$).

(ii) The Z characteristic for the $F \rightarrow S$ transition [Fig. 2(a)] *together with* the Z characteristic for the $S \rightarrow J$ transition [Fig. 4(a)] form a *double* Z-shaped speed-density characteristic related to the $F \rightarrow S \rightarrow J$ transitions [Fig. 4(b)]. At the chosen model parameters $q_{out}<q_{th}$. In the range $q_{out} \leqslant q_{in} < q_{th}$, MSP either dissolves or transforms into a wide moving jam [34]. At $q_{th} \leqslant q_{in} < q_{max}$ both MSP and wide moving jams can occur and exist.

(iii) For a road with the on-ramp, there is also a Z-shaped speed-flow characteristic [Fig. 4(c)] which consists of states $v_{syn}^{(B)}$ (dashed 2D region), of the branch $v_{cr,SJ}^{(B)}$ which determines the critical amplitude of local perturbations $\Delta v_{cr,SJ}^{(B)} = v_{syn}^{(B)} - v_{cr,SJ}^{(B)}$ [33], and of the line $v_{min}=0$. The Z characteristic shows the metastability of synchronized flow $v_{syn}^{(B)}$ and the related *spontaneous* nucleation effect leading to the $S \rightarrow J$ transition and to the GP emergence. This occurs after a random delay time T_{SJ} in the range $q_{on}^{(th,SJ)} \leqslant q_{on} \leqslant q_{on}^{(cr,SJ)}$ if the perturbation amplitude exceeds $\Delta v_{cr}^{(SJ)}$ (down arrow); otherwise, the perturbation decays (up arrow). The exact point $q_{on}=q_{on}^{(cr,SJ)}$ cannot be found: At $q_{on} \approx q_{on}^{(cr,SJ)}$ $T_{SJ} \approx$ 10 min, which is comparable with the duration of the $S \rightarrow J$ transition (dashed parts of $v_{syn}^{(B)}$ and $v_{cr,SJ}^{(B)}$).

(iv) The Z characteristic for the $F \rightarrow S$ transition [Fig. 3(a)] *together with* the Z characteristic for the $S \rightarrow J$ transition [Fig. 4(c)] form a *double* Z-shape speed-density characteristic at the on-ramp [Fig. 4(d)] related to the spontaneous $F \rightarrow S \rightarrow J$ transitions and the GP emergence. This is due to two different spontaneous nucleation effects in free flow $v_{free}^{(B)}$ and synchronized flow $v_{syn}^{(B)}$, respectively. The double Z characteristic depends on q_{in} [curves 1–4, Fig. 4(e)]. At $q_{in} \geqslant q_{th}$ and $q_{on}=0$ there are gaps Δ_S and Δ_J in the speed on the double Z characteristic [Fig. 4(e), curve 1]. The gap Δ_S is linked to the speed gap between states $v_{cr}^{(FS)}$ and v_{syn} in Fig. 2(a). The gap Δ_J is linked to the speed gap between states $v_{cr}^{(SJ)}$ and $v_{min}=0$ in Fig. 4(a). At $q_{in}<q_{th}$ the gaps do not appear [curves 3 and 4, Fig. 4(e)] [35].

V. STRONG AND WEAK CONGESTION AT ISOLATED ON-RAMP

A. Diagram of congested patterns: Synchronized flow patterns

The diagram of different congested patterns, i.e., the regions of the *spontaneous* occurrence of the patterns in the flow-flow plane whose coordinates are q_{in} and q_{on}, is qualitatively similar to the diagram postulated in Ref. [17] and found out in Ref. [12] (Fig. 5). Different SP [Figs. 5(c)–5(e) and 6] occur between the boundaries $F_S^{(B)}$ and $S_J^{(B)}$ in Fig. 5(a). $F_S^{(B)}$ and $S_J^{(B)}$ are related to the *spontaneous* $F \rightarrow S$ and $S \rightarrow J$ transitions, respectively. The downstream front of the widening SP (WSP) is fixed at the on-ramp [Fig. 5(c)]. The upstream front of WSP is continuously widening upstream. WSP occurs above the boundary W in Fig. 5(a). Below the boundary W the localized SP (LSP) occurs. As in WSP, the downstream front of LSP is fixed at the on-ramp. However, the upstream front of LSP is localized at a some distance,

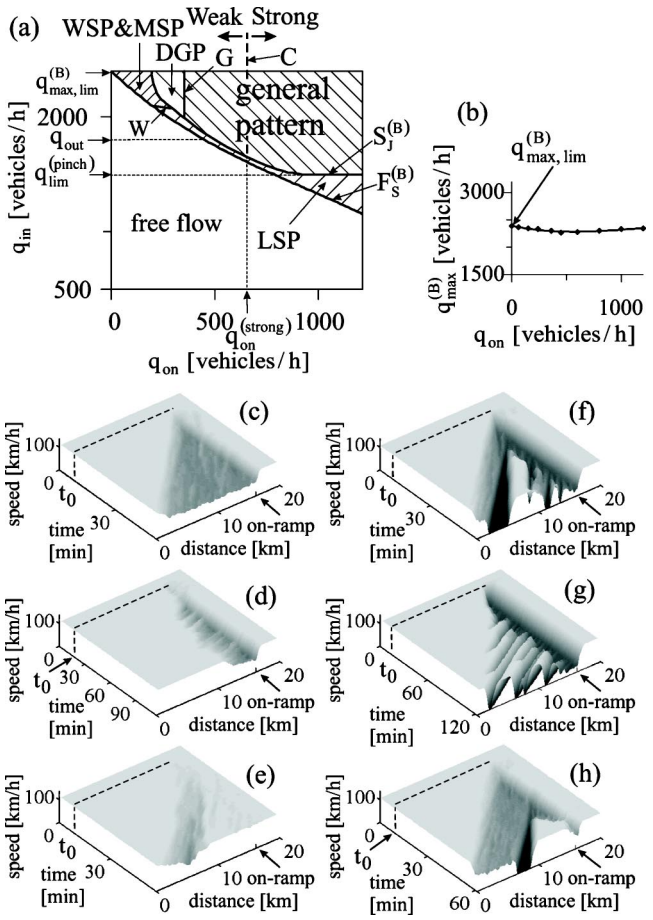

FIG. 5. Diagram of congested patterns at the isolated on-ramp (a), the capacity in free flow at the on-ramp (b) and congested patterns (c)–(h) related to (a): (c)–(e) SP and (f)–(h) GP. (c) The widening SP (WSP), (d) the localized SP (LSP), (e) the moving SP (MSP), (f) GP at $q_{in}>q_{out}$, (g) GP at $q_{in}<q_{out}$, (h) the dissolving GP (DGP). In (c)–(h) the flow rates (q_{on},q_{in}) are (c) (260,2280), (d) (310,1945), (e) (35,2307), (f) (500,2250), (g) (1200,1658), and (h) (250,2250) vehicles/h. In (a) the criteria for the boundaries $F_S^{(B)}$ and $S_J^{(B)}$ are the same as used and explained in Ref. [28]. $q_{max,lim}^{(B)} \approx 2400$ vehicles/h. The on-ramp inflow is switched on at $t=t_0=8$ min.

L_{LSP}, upstream of the on-ramp [Fig. 5(d)]. At higher q_{in} and a very low q_{on} MSP can occur [Fig. 5(e)] rather than WSP. We have found the following features of SP.

(i) The flow rate in SP is often only slightly lower than the initial flow rate in free flow [Figs. 6(a,b,e), right].

(ii) The density in SP can often be nearly the same or even lower than $\rho_{max}^{(free)}$ [Figs. 6(c,f,g)].

(iii) The width of LSP, L_{LSP}, i.e., the distance of the upstream LSP front from the on-ramp depends on time and it can show large amplitude complex oscillations [Fig. 6(d) where 2 km$\leqslant L_{LSP} \leqslant$4 km].

(iv) The mean width of LSP, $L_{LSP}^{(mean)}$, can depend strongly on q_{in} and q_{on}: $L_{LSP}^{(mean)}$ can be changed between 0.5 km and 10 km.

(v) At q_{on} and q_{in} which are related to a vicinity of the boundary $F_S^{(B)}$ sometimes regions of free flow occur inside WSP and LSP. These patterns are called ''SP with alterna-

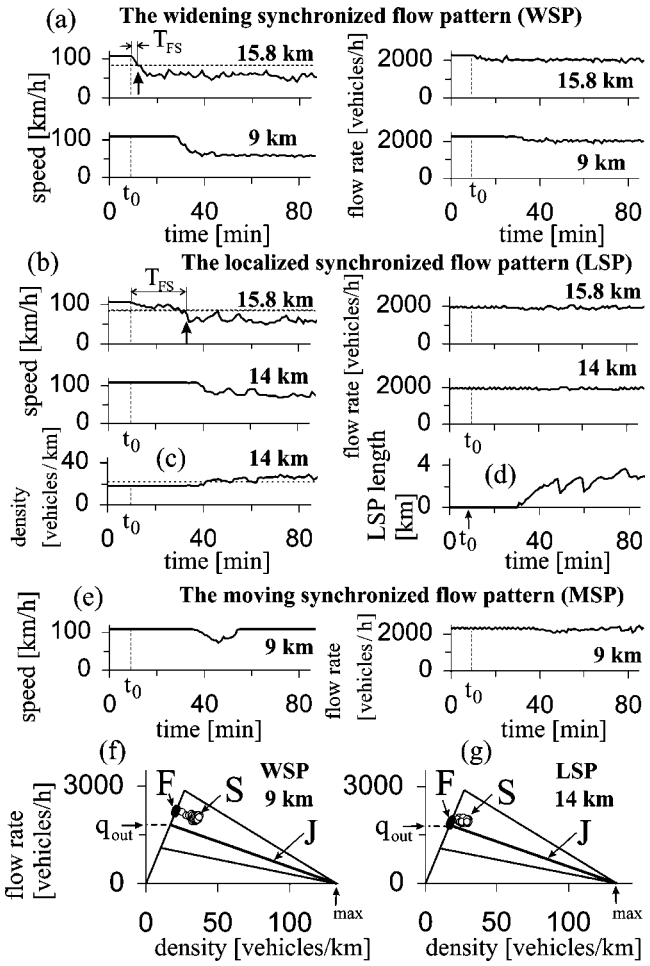

FIG. 6. Synchronized flow patterns (SP). (a),(b),(e) The vehicle speed (left) and the flow rate (right) in SP shown in Figs. 5(c)–5(e), respectively. (a) WSP, (b) LSP, (c) the vehicle density in LSP over time, (d) the length of LSP over time, (e) MSP. The data in the flow-density plane in (f) are shown for WSP (a) and in (g) for LSP (b). The dashed line in (c) is related to $\rho_{\max}^{(\mathrm{free})}\approx 22.5$ vehicles/km. $t_0=8$ min.

tions of free and synchronized flow'' (ASP) [17,37]. The appearance of ASP may be explained by the $S\to F$ transition inside SP. There are two different cases.

(1) Free flow regions occur at some distance upstream of the on-ramp. At the bottleneck synchronized flow is self-maintained. Often regions of free flow appear only for a finite time, i.e., free flow is replaced by synchronized flow over time.

(2) Free flow regions occur for short time intervals randomly at the on-ramp. The free flow replaces a part of synchronized flow upstream. Here there is an analogy with the occurrence of MSP. However, in contrast to the case of MSP, free flow returns at the on-ramp for much shorter time intervals than the duration of synchronized flow at the on-ramp.

(vi) In the model, freeway capacity in free flow at the on-ramp $q_{\max}^{(\mathrm{B})}=q_{\mathrm{sum}}|_{F_S^{(\mathrm{B})}}$ which is related to the boundary $F_S^{(\mathrm{B})}$ (here $q_{\mathrm{sum}}=q_{\mathrm{on}}+q_{\mathrm{in}}$ is the flow rate downstream of the on-ramp in free flow at the bottleneck) is a weak function of q_{on} with a minimum point [Fig. 5(b)]. It must be stressed that the *limit* capacity $q_{\max}^{(\mathrm{B})}|_{q_{\mathrm{on}}\to 0}=q_{\max,\mathrm{lim}}^{(\mathrm{B})}$ (but $q_{\mathrm{on}}\neq 0$) which is shown in Figs. 5(a,b) is *lower* than the maximal flow rate $q_{\max}$ for a homogeneous road [see Eq. (13) and Fig. 2(a)]: $q_{\max,\mathrm{lim}}^{(\mathrm{B})}<q_{\max}$ [27]. To understand this result, note that at a very small flow rate to the on-ramp ($q_{\mathrm{on}}\approx 1-2$ vehicles/h) *single* vehicles squeezing from the on-ramp to the main road cause *random local short-time* perturbations on the main road in the on-ramp vicinity at high enough flow rate q_{in}. These perturbations lead to the *spontaneous* speed breakdown (the $F\to S$ transition) at the on-ramp even when $q_{\mathrm{on}}\to 0$ (but $q_{\mathrm{on}}\neq 0$) at the lower flow rate $q_{\mathrm{sum}}=q_{\max,\mathrm{lim}}^{(\mathrm{B})}$ than the maximal flow rate $q_{\max}$ on the homogeneous road.

(vii) At $q_{\mathrm{on}}\to 0$, i.e., in the vicinity of $q_{\mathrm{sum}}=q_{\max,\mathrm{lim}}^{(\mathrm{B})}$ the local short-time fluctuations [item (vi)] can play the role of a nucleus for the spontaneous MSP emergence at the on-ramp. After the MSP has emerged, the MSP comes off the on-ramp and begins to propagate as an independent localized structure [Fig. 5(e)] as on the homogeneous road [Fig. 2(a)] [36].

B. Features of GP

In the diagram [Fig. 5(a)], right of the boundary G and right of the boundary $S_{\mathrm{J}}^{(\mathrm{B})}$ GP occurs where the pinch region continuously exists and a sequence of wide moving jams emerges [Figs. 5(f,g)]. Right of the boundary $S_{\mathrm{J}}^{(\mathrm{B})}$ and left of the boundary G dissolving GP (DGP) occurs [Fig. 5(h)] [12]. We found the following features of GP and DGP. (i) In the pinch region of GP narrow moving jams spontaneously emerge and grow [Figs. 7(a,b), $x=15.8$ km and $x=14.5$ km]. In agreement with Ref. [25], points related to the pinch region lie above the line J in the flow-density plane [Fig. 7(c)]. After a narrow moving jam has transformed into a wide moving jam, this jam can suppress the growing of the downstream narrow moving jam [Fig. 7(a), $x=13$ km]. In this case, the mean time distance between narrow moving jams emerging in the pinch region T_{J} is noticeably lower than the mean time distance between wide moving jams, $T_{\mathrm{J}}^{(\mathrm{wide})}$ (in Fig. 7(a), $T_{\mathrm{J}}\approx 6$ min at $x=14.5$ km and $T_{\mathrm{J}}^{(\mathrm{wide})}\approx 30$ min at $x=8$ km). However, the suppressing effect occurs only if the narrow moving jam is close to the downstream wide moving jam front.

(ii) Thus, if the distance between the initial narrow moving jams emerging in the pinch region R_{narrow} is low, i.e., the frequency of the narrow jam emergence in the pinch region f_{narrow} is high (at the model parameters under consideration $R_{\mathrm{narrow}}<2.5$ km and the related frequency $f_{\mathrm{narrow}}>0.1\,\mathrm{min}^{-1}$) then only some of the initial narrow moving jams transform into wide moving jams: The other narrow moving jams either dissolve or merge with other moving jams. If in contrast, R_{narrow} is relatively high, i.e., f_{narrow} is low then almost each narrow jam transforms into a wide moving jam [Fig. 8(a)].

(iii) In DGP, after a wide moving jam has been formed this jam suppresses the pinch region in the initial GP [12,17]. As a result, either free flow or a LSP appears at the on-ramp while the wide moving jam moves upstream away from the on-ramp [Fig. 5(h)]. In a comparison with Ref. [12], the boundary G Fig. 5(a) does not intersect the boundary $S_{\mathrm{J}}^{(\mathrm{B})}$ at

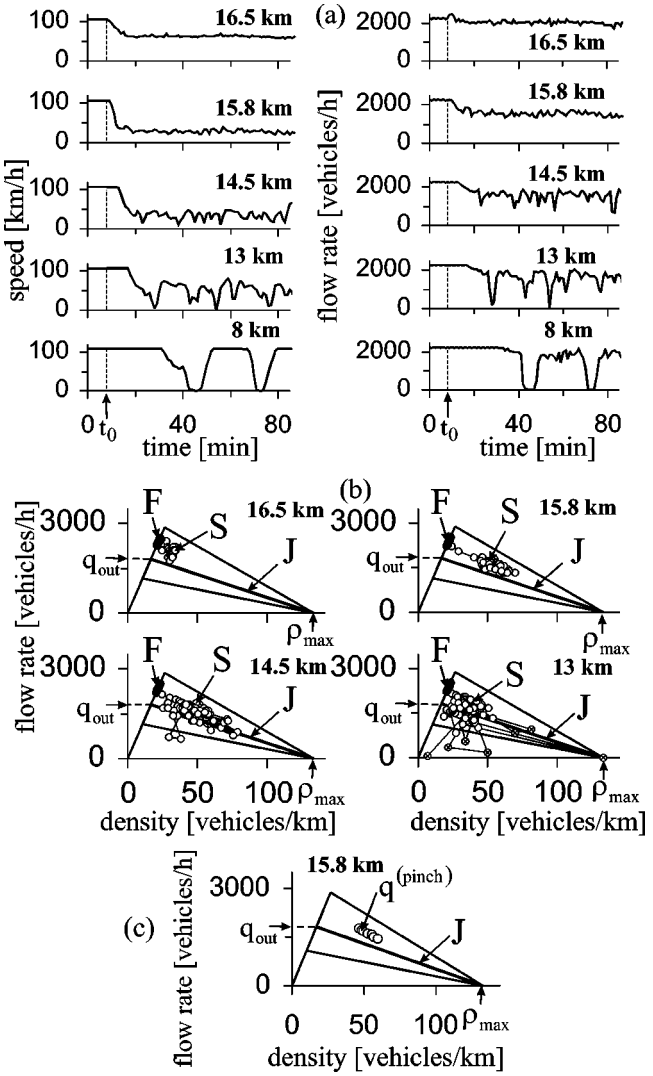

FIG. 7. The general pattern (GP). (a) The vehicle speed (left) and the flow rate (right), (b) the corresponding data in the flow-density plane for the GP shown in Fig. 5(f). (c) Points in the flow-density plane related to synchronized flow in the pinch region of GP. (a),(b) 1-min averaged data of virtual detectors. In (b) black points and circles are related to free flow (F) and to synchronized flow (S), respectively. With the symbol $\otimes$ the data related to moving jams are designated. $t_0=8$ min.

the point $q_{in}=q_{out}$. This is due to hysteresis effects of the GP formation (Sec. VI).

(iv) If $q_{in}>q_{out}$ then the width of the most upstream wide moving jam is a gradually increasing function of time [Fig. 5(f)]. If in contrast $q_{in}<q_{out}$, the most upstream wide moving jam gradually dissolves [see the dissolving of the first three wide moving jams in GP in Fig. 5(g)]. Even if $q_{in}<q_{out}$, the region of wide moving jams is expanded in the downstream direction.

(v) In a lot of cases GP consists of the pinch region in synchronized flow and the region of wide moving jams. Then, the width of synchronized flow in GP, L_{syn} is the distance between the downstream front of synchronized flow in GP (located at the on-ramp) and the upstream boundary of the pinch region. This L_{syn} is spatially limited.

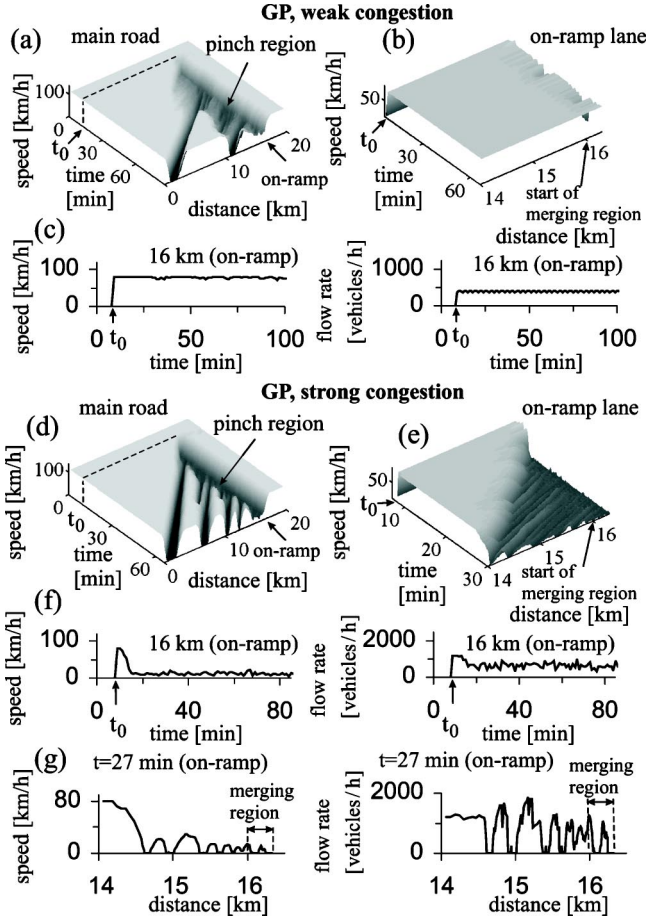

FIG. 8. Weak congestion and strong congestion in GP at the on-ramp. (a), (b) The speed in space and time in GP under weak congestion on the main road (a) and on the on-ramp lane (b). (c) The vehicle speed (left) and the flow rate (right) related to (b). The speed in space and time in GP under strong congestion on the main road (d) and on the on-ramp lane (e). (f) The vehicle speed (left) and the flow rate (right) related to (e). (g) Spatial distributions of the speed (left) and the flow rate (right) along the on-ramp lane at $t=27$ min related to (e). $t_0=8$ min. $q_{in}=2000$ vehicles/h, q_{on} has the following values: 400 (a)–(c), 1200 (d)–(g) vehicles/h.

(vi) In contrast, there is a case where L_{syn} in GP can be a continuous increasing function over time. This can occur above the boundary W in Fig. 5(a) when q_{on} is gradually increased. First WSP occurs, then DGP appears, and finally GP is realized. In this case, the pinch region and the moving jam emergence in DGP and GP occur inside the initial WSP.It can occur that the upstream front of the initial WSP propagates upstream quicker than the upstream front of the wide moving jam [Fig. 5(f)], i.e., the upstream front of the whole GP can be determined by the upstream front of the initial WSP rather than of the most upstream wide moving jam.

C. Weak and strong congestion in GP

It has been found that there is a special flow rate to the on-ramp

$$q_{on}=q_{on}^{(strong)}. \tag{14}$$

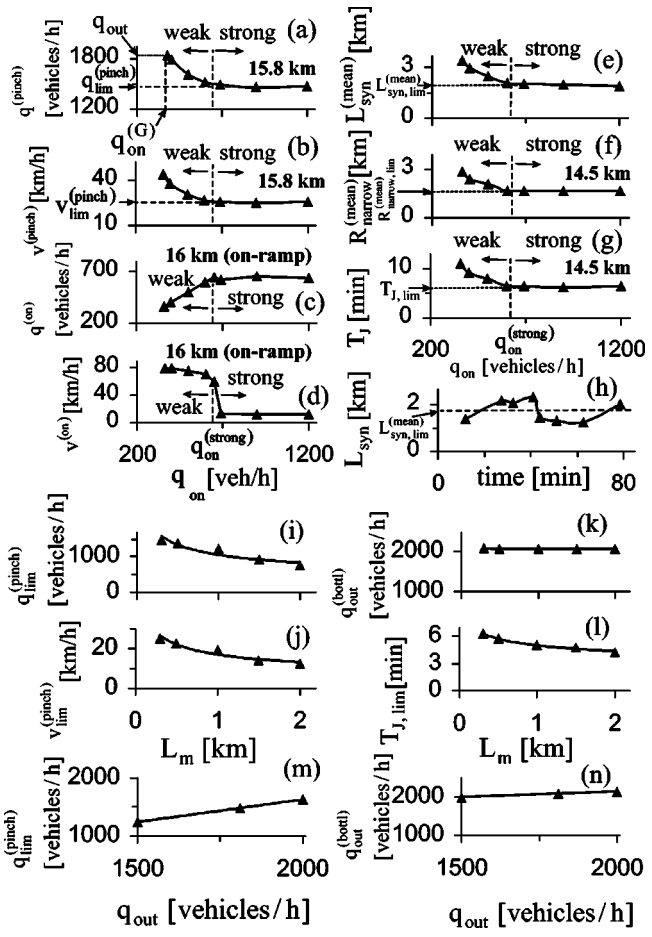

FIG. 9. Parameters of GP under weak and strong congestion conditions: $q^{(\mathrm{pinch})}$ (a), $v^{(\mathrm{pinch})}$ (b), $q^{(\mathrm{on})}$ (c), $v^{(\mathrm{on})}$ (d), $L_{\mathrm{syn}}^{(\mathrm{mean})}$ (e), $R_{\mathrm{narrow}}^{(\mathrm{mean})}$ (f), and T_{J} (g) as functions of q_{on}. (h) L_{syn} as a function of time. $q_{\mathrm{lim}}^{(\mathrm{pinch})}$ (i), $v_{\mathrm{lim}}^{(\mathrm{pinch})}$ (j), $q_{\mathrm{out}}^{(\mathrm{bottle})}$ (k), and T_{J} (l) under strong congestion as functions of L_{m} [Fig. 1(c)]. (m),(n) $q_{\mathrm{lim}}^{(\mathrm{pinch})}$ (m) and $q_{\mathrm{out}}^{(\mathrm{bottle})}$ (n) vs q_{out}. In (a)–(d) the interval of data averaging is 60 min. In (m),(n) q_{out} is changed by multiplying the function $p_0(v_{\mathrm{n}})$ (Sec. II C) by a numerical factor.

If the condition $q_{\mathrm{on}}<q_{\mathrm{on}}^{(\mathrm{strong})}$ is fulfilled then the averaged flow rate in the pinch region of GP, $q^{(\mathrm{pinch})}$, is a decreasing function of traffic demand, i.e., of q_{on} [Fig. 9(a)]. The same behavior shows the average vehicle speed $v^{(\mathrm{pinch})}$ in the pinch region [Fig. 9(b)]. The interval of the averaging of $q^{(\mathrm{pinch})}$ and $v^{(\mathrm{pinch})}$ is chosen to be higher than T_{J}. The mean width of the pinch region, $L_{\mathrm{syn}}^{(\mathrm{mean})}$, the mean space distance between narrow moving jams in the pinch region, $R_{\mathrm{narrow}}^{(\mathrm{mean})}$, and the related value T_{J} are decreasing functions of q_{on} [Figs. 9(e)–9(g)]. Consequently, the frequency f_{narrow} is an increasing function of q_{on}. Following Ref. [17], this case is called "weak congestion" [GP in Fig. 8(a)–8(c)].

In weak congestion, the flow rate of the vehicles which are really squeezing to the main road from the on-ramp, $q^{(\mathrm{on})}$ is nearly equal to q_{on} [Fig. 9(c)]. Also the vehicle speed on the on-ramp lane leading to the merging region of the on-ramp $v^{(\mathrm{on})}\approx v_{\mathrm{free\ on}}$ [Fig. 9(d); see also Figs. 8(b,c); $v^{(\mathrm{on})}$ and $q^{(\mathrm{on})}$ are related to figures which have been marked "16 km (on-ramp)"].

In contrast, when $q_{\mathrm{on}}\geqslant q_{\mathrm{on}}^{(\mathrm{strong})}$, then the mean characteristics of the pinch region and of the narrow moving jam emergence do not depend on traffic demand, i.e., on q_{on}, anymore. The mean values $q^{(\mathrm{pinch})}$, $v^{(\mathrm{pinch})}$, $L_{\mathrm{syn}}^{(\mathrm{mean})}$, $R_{\mathrm{narrow}}^{(\mathrm{mean})}$, and T_{J} as functions of $q^{(\mathrm{on})}$ are saturated and reach some limit (minimum) values: $q^{(\mathrm{pinch})}=q_{\mathrm{lim}}^{(\mathrm{pinch})}$, $v^{(\mathrm{pinch})}=v_{\mathrm{lim}}^{(\mathrm{pinch})}$, $L_{\mathrm{syn}}^{(\mathrm{mean})}=L_{\mathrm{syn,\,lim}}^{(\mathrm{mean})}$, $R_{\mathrm{narrow}}^{(\mathrm{mean})}=R_{\mathrm{narrow,\,lim}}^{(\mathrm{mean})}$, $T_{\mathrm{J}}=T_{\mathrm{J,\,lim}}$ [Figs. 9(a)–9(g)], correspondingly f_{narrow} reaches a maximum $f_{\mathrm{narrow}}=f_{\mathrm{max}}=1/T_{\mathrm{J,\,lim}}$. In empirical observations, this case has been called "strong congestion" [17]. Note that $L_{\mathrm{syn}}(t)$ changes noticeably near $L_{\mathrm{syn\,lim}}^{(\mathrm{mean})}$ [Fig. 9(h)]. This is because transformations of different narrow moving jams into wide moving jams occur at different locations.

If q_{on} noticeably exceeds $q_{\mathrm{on}}^{(\mathrm{strong})}$ then there is a saturation effect in the boundary $S_{\mathrm{J}}^{(\mathrm{B})}$. This effect is realized at a flow rate $q_{\mathrm{in}}=q_{\mathrm{lim}}^{(\mathrm{pinch})}$ considerably lower than q_{out} [Fig. 5(a)]. Only when $q_{\mathrm{in}}<q_{\mathrm{lim}}^{(\mathrm{pinch})}$ GP dissolves (see also Sec. VI). Thus, GP can exist only if

$$q_{\mathrm{in}}\geqslant q_{\mathrm{lim}}^{(\mathrm{pinch})}. \tag{15}$$

There is a limit flow rate $q_{\mathrm{on}}=q_{\mathrm{on}}^{(\mathrm{G})}$ [see $q^{(\mathrm{pinch})}(q_{\mathrm{on}})$ in Fig. 9(a)] which gives the boundary G in the diagram [Fig. 5(a)] and it is related to the condition

$$q^{(\mathrm{pinch})}|_{q_{\mathrm{on}}=q_{\mathrm{on}}^{(\mathrm{G})}}=q_{\mathrm{out}}. \tag{16}$$

GP exists only if

$$q^{(\mathrm{pinch})}<q_{\mathrm{out}} \tag{17}$$

at

$$q_{\mathrm{on}}>q_{\mathrm{on}}^{(\mathrm{G})}. \tag{18}$$

At $q_{\mathrm{on}}<q_{\mathrm{on}}^{(\mathrm{G})}$ ($q^{(\mathrm{pinch})}>q_{\mathrm{out}}$) the pinch region dissolves and GP cannot exist. Indeed, the average flow rate in the pinch region $q^{(\mathrm{pinch})}$ is equal to the average flow rate in the region of wide moving jams. The latter cannot exceed q_{out}.

D. Strong congestion features

The following features of strong congestion in GP have been found.

(i) The mean time distance between the downstream fronts of wide moving jams $T_{\mathrm{J}}^{(\mathrm{wide})}$ under strong congestion is lower [Fig. 8(d)] than under weak congestion [Fig. 8(a)]. Because the mean time-width of wide moving jams, τ_{J}, is similar in the both cases [$\tau_{\mathrm{J}}\approx 2$ min in Figs. 8(d) and 8(a)] the mean time intervals between wide moving jams, $T_{\mathrm{int}}^{(\mathrm{wide})}=T_{\mathrm{J}}^{(\mathrm{wide})}-\tau_{\mathrm{J}}$, are very different: The value $T_{\mathrm{int}}^{(\mathrm{wide})}$ in strong congestion can be considerably lower than under weak congestion.

(ii) The limit characteristics $q_{\mathrm{lim}}^{(\mathrm{pinch})}$, $v_{\mathrm{lim}}^{(\mathrm{pinch})}$, and $T_{\mathrm{J,\,lim}}$ are decreasing functions of the length of merging region L_{m} whereas the discharge flow rate $q_{\mathrm{out}}^{(\mathrm{bottle})}$ is independent of L_{m} [Figs. 9(i)–9(l)]. $q_{\mathrm{out}}^{(\mathrm{bottle})}$ is the flow rate in the outflow from a congested pattern at the bottleneck determined downstream of the on-ramp where free flow is realized.

(iii) There is a correlation between features of wide moving jams and $q^{(\text{pinch})}$. The averaged flow rate within the region of wide moving jams is $q_{\text{J}}^{(\text{wide})}=q_{\text{out}}(T_{\text{int}}^{(\text{wide})}/T_{\text{J}}^{(\text{wide})})$ provided the interval of the averaging is much larger than $T_{\text{J}}^{(\text{wide})}$. This flow determines $q^{(\text{pinch})}$: $q^{(\text{pinch})}=q_{\text{J}}^{(\text{wide})}$. Under the strong congestion condition $q^{(\text{pinch})}=q_{\text{lim}}^{(\text{pinch})}$, i.e.,

$$q_{\text{lim}}^{(\text{pinch})}/q_{\text{out}}=T_{\text{int}}^{(\text{wide})}/T_{\text{J}}^{(\text{wide})}<1. \tag{19}$$

For the GP in Fig. 8(d) $q_{\text{lim}}^{(\text{pinch})}=1470$ vehicles/h, therefore $q_{\text{out}}/q_{\text{lim}}^{(\text{pinch})}=1.23$.

(iv) The higher q_{out} is, the higher is $q_{\text{lim}}^{(\text{pinch})}$. This dependence is nearly a liner one [Fig. 9(m)]. $q_{\text{out}}^{(\text{bottle})}$ also increases with q_{out} [Fig. 9(n)]. However, this dependence is much more weaker than the former one. Note that the condition $q_{\text{in}}+q_{\text{on}}>q_{\text{out}}^{(\text{bottle})}$ is fulfilled. Therefore $q_{\text{out}}^{(\text{bottle})}$ gives the *congested pattern capacity* [27,28].

(v) The maximal highway capacity in free flow, $q_{\text{max,lim}}^{(\text{B})}$, is considerable higher than $q_{\text{lim}}^{(\text{pinch})}$: $q_{\text{max,lim}}^{(\text{B})}/q_{\text{lim}}^{(\text{pinch})}=1.63$.

(vi) GP under strong congestion occurs also on the on-ramp lane. This occurs at a little bit *higher* q_{on} than $q_{\text{on}}^{(\text{strong})}$ (14) [Fig. 9(d)]. In this case, $q^{(\text{on})}$ is saturated to a limit value [Fig. 9(c)]. A further increase in q_{on} at the beginning of the on-ramp lane [at $x=x_{on}^{(\text{b})}=14$ km, Fig. 1(c)] does not lead to the increase in $q^{(\text{on})}$ anymore. The dependence of $v^{(\text{on})}$ on q_{on} possesses a sharp decrease when the transition to strong congestion occurs at the on-ramp lane [Fig. 9(d)]. If q_{on} is initially set much higher than $q_{\text{on}}^{(\text{strong})}$, the sharp decrease in the speed and the flow rate occurs first in the vicinity of the merging region of the on-ramp [Fig. 8(f)]. Then the front separating lower and higher speeds (and the flow rates) on the on-ramp lane propagates upstream up to $x=x_{on}^{(\text{b})}$ [Fig. 8(e)]. Consequently, the pinch region is formed on the on-ramp lane. Growing narrow moving jams emerge in this pinch region and finally some of these jams transform into wide moving jams [Fig. 8(g)]. In contrast to the main road, where narrow moving jams transform into wide moving jams at distances about 1.5–2 km downstream from the on-ramp, some of wide moving jams on the on-ramp lane already occur within the merging region. Some of vehicles moving under the strong congestion condition on the on-ramp lane have to stop before they can enter the main road.

E. Evolution of congested patterns at on-ramps

To simulate the pattern evolution, empirical $q_{\text{on}}(t)$ and $q_{\text{in}}(t)$ from Ref. [17] have been used [Fig. 10(a)]. In accordance with Ref. [17] we found that during the increase in $q_{\text{on}}(t)$ and $q_{\text{in}}(t)$ first the $F\rightarrow S$ transition occurs spontaneously at the on-ramp at $t=6{:}12$ and then synchronized flow appears upstream of the on-ramp [Figs. 10(b) and 10(c) at $t=6{:}20$]. Because $q_{\text{on}}(t)$ further increases sharply [Fig. 10(a)] the pinch effect is realized in this synchronized flow leading to the GP emergence at $t=6{:}36$ [Fig. 10(b)]. The maximum of $q_{\text{on}}(t)$ exceeds $q_{\text{on}}^{(\text{strong})}$ noticeably. Thus, as in empirical Fig. 11 in Ref. [17], in GP strong congestion is realized.

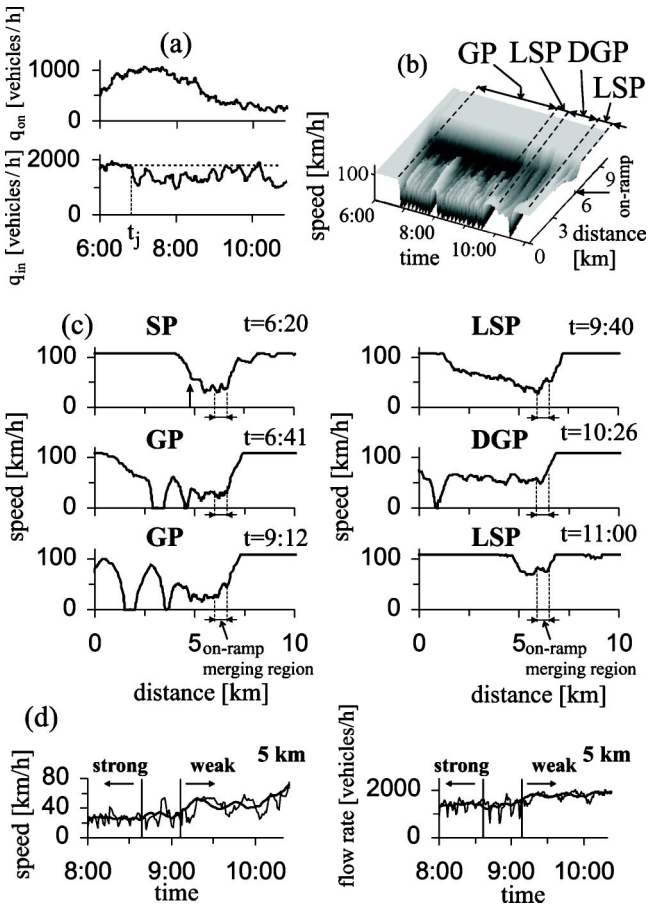

FIG. 10. Evolution of congested patterns at the on-ramp. (a) $q_{\text{on}}(t)$ (top) and $q_{\text{in}}(t)$ (bottom) related to the effective flow rate "off-on" and to the flow rate at the detectors D1, respectively taken from empirical data in Fig. 12 of Ref. [17] (the empirical data have been multiplied by numerical factors: 0.67 for q_{on} and 0.42 for q_{in}). (b) The pattern evolution. (c) The speed distributions. (d) The speed (left) and the flow rate (right) at a virtual detector in the pinch region of GP. t_{j} in (a) is the time moment at which the first wide moving jam reaches the detectors D1 in the empirical data in Fig. 12 from [17]. At $t\geqslant t_{\text{j}}$ q_{in} is chosen as $q_{\text{in}}=q_{\text{in}}(t_{\text{j}})$. In (d) 1 min (thin lines) and 10 min (thick lines) averaged data are presented. The on-ramp parameters [Fig. 1(c)] are $x_{\text{on}}=6$ km, $L_{\text{m}}=600$ m, $L_{\text{r}}=300$ m.

Over time q_{on} begins to decrease [Fig. 10(a)]. As in empirical Fig. 18(a) [17], the transition to weak congestion in GP occurs [Fig. 10(d)]. During further decrease in q_{on} [Fig. 10(a)] GP transforms into LSP at $t=9{:}30$ [compare Fig. 10(b) and empirical Fig. 19(a) (left) [17]]. The vehicle speed in LSP is higher than in the pinch region of GP. During the further evolution of synchronized flow in LSP, the pinch effect in this synchronized flow occurs where a wide moving jam emerges: At $t=10{:}15$ LSP transforms into DGP [Figs. 10(b,c)]. As in Fig. 19 in Ref. [17], this pinch effect occurs at some distance (about 3 km) upstream from the on-ramp. After the wide moving jam of DGP is far away from the on-ramp, LSP appears at $t>10{:}35$. LSP disappears and free flow occurs at the on-ramp when q_{on} further decreases.

VI. HYSTERESIS AND NUCLEATION EFFECTS BY PATTERN FORMATION AT ON-RAMPS

A. Threshold boundary for SP

The boundary $F_S^{(B)}$ [Fig. 5(a)] corresponds to the spontaneous $F \to S$ transition which occurs at the on-ramp for a given time interval ($T_0 = 30$ min) with the probability $P_{FS} \approx 1$. Left of $F_S^{(B)}$, i.e., in free flow, nevertheless, the $F \to S$ transition at the on-ramp can spontaneously also occur with the probability $P_{FS} < 1$ [28].

This result is explained by the Z-shaped function in Fig. 3(a). For a given $q_{in} < q_{th}$ the threshold flow rate $q_{on}^{(th)}$ at this Z characteristic is related to a point on a threshold boundary $F_{th}^{(B)}$ in Fig. 11(a). By the approaching $F_{th}^{(B)}$ the probability of the $F \to S$ transition $P_{FS} \to 0$. Thus, left of $F_{th}^{(B)}$ in the plane (q_{on}, q_{in}) no SP can exist for a long time. Between the boundaries $F_{th}^{(B)}$ and $F_S^{(B)}$ [dashed region in Fig. 11(a)] dependent on initial conditions either free flow or one of SP can exist. This region in the diagram is a metastable region of free flow with respect to the $F \to S$ transition and to the SP formation.

SP can later exist for a long time (>50 min). In some cases depending on q_{on} and q_{in} the SP can either dissolve or transform into GP. There is the hysteresis effect related to the metastable states [Figs. 12(a,b)]. After SP has first spontaneously occurred right of the boundary $F_S^{(B)}$, the SP can also further exist when one of the flow rates q_{on} or q_{in} (or both of them) decrease and the point (q_{on}, q_{in}) is between $F_S^{(B)}$ and $F_{th}^{(B)}$. SP disappears near the threshold boundary $F_{th}^{(B)}$. At $q_{on} \to 0$ the flow rate $q_{in} \to q_{th}$ (Sec. III) at the threshold boundary $F_{th}^{(B)}$.

In some interval of low q_{on}, the flow rate q_{in} at $F_{th}^{(B)}$ remains nearly a constant: $q_{in} \approx q_{th}$. There is a boundary [the vertical boundary marked M in Fig. 11(a)] which separates the region of MSP (left of M) and WSP (right). Left of the boundary M and between the boundaries $F_{th}^{(B)}$ and $F_S^{(B)}$ MSP can either spontaneously occur [Sec. V A, item (vii)] or be excited by the applying of a short-time perturbation of q_{on}. At lower q_{on} single MSP occurs [Fig. 5(e)]. If q_{on} increases (q_{on} is still related to a point in the diagram left of the boundary M) first WSP spontaneously occurs. However, inside this WSP local regions of free flow spatially alternating with initial synchronized flow spontaneously appear over time. This ASP finally transforms into a sequence of MSP. In contrast, between the boundaries $F_{th}^{(B)}$ and $F_S^{(B)}$ and right of the boundary M (and also above the boundary W) rather than MSP, WSP can be excited or spontaneously occur.

B. Threshold boundary for GP

The hysteresis effect and metastable states appear also by the GP formation. One finds the following.

(i) At the boundary $S_J^{(B)}$ the $S \to J$ transition occurs for a given time interval [$T_0 = 60$ min in Figs. 5(a) and 11(b)] with the probability $P_{SJ} \approx 1$. However, left of $S_J^{(B)}$ the $S \to J$ transition can also occur with the probability $P_{SJ} < 1$. This result is linked to the Z-shaped characteristic for the $S \to J$ transition [Fig. 4(c)]. For a given $q_{in} < q_{out}$ the flow rate $q_{on}^{(th,\,SJ)}$ at this Z characteristic is related to a point at a threshold boundary $J_S^{(B)}$ in the diagram. Below and left of the boundary $J_s^{(B)}$, the probability $P_{SJ} = 0$ [Fig. 11(b)]. Between $S_J^{(B)}$ and $J_S^{(B)}$ [dashed regions in Fig. 11(b)] free flow and SP are *metastable* with respect to the GP formation, i.e., GP can either spontaneously occur during a high enough time interval or be excited.

(ii) If a GP has spontaneously occurred right of $S_J^{(B)}$, this GP can further exist if one of the flow rates q_{in} and q_{on} (or both) decreases so that the point in the diagram is displaced into the region between the boundaries $S_J^{(B)}$ and $J_S^{(B)}$. Below and left of $J_S^{(B)}$ GP dissolves.

(iii) The boundary $J_S^{(B)}$ consists of a horizontal line

$$q_{in} = q_{out} \quad \text{at} \quad q_{on} \leqslant q_{on}^{(G)} \tag{20}$$

and a curve determined by the condition

$$q_{in} = q^{(pinch)}(q_{on}) \quad \text{at} \quad q_{on} > q_{on}^{(G)}, \tag{21}$$

where the dependence $q^{(pinch)}(q_{on})$ is shown in Fig. 9(a). These two branches of $J_S^{(B)}$ merge at a point $q_{on} = q_{on}^{(G)}$ (16). At this point, the boundary $J_S^{(B)}$ intersects the boundary G which separates DGP and GP in the diagram [Fig. 11(b)]. Thus, between $J_S^{(B)}$ and $S_J^{(B)}$ and *left* of the boundary G metastable states exist where DGP can spontaneously occur and be excited. In contrast, the region between $J_S^{(B)}$ and $S_J^{(B)}$ and *right* of G is the metastable region with respect to the GP formation.

(iv) To explain the boundary $J_S^{(B)}$, recall that if condition (18) is not valid, i.e., at $q_{on} < q_{on}^{(G)}$ the pinch region and GP cannot continuously exist. However, if $q_{in} > q_{out}$ a *single* wide moving jam, i.e., DGP, can exist. In contrast, at $q_{in} < q_{out}$ the single wide moving jam cannot be formed and DGP cannot appear. This explains Eq. (20). At $q_{on} > q_{on}^{(G)}$ (18), GP can occur if the inflow in GP q_{in} is higher than $q^{(pinch)}$. In contrast, at $q_{in} < q^{(pinch)}$ the pinch region and GP dissolve. Thus, condition (21) indeed determines $J_S^{(B)}$ at $q_{on} \geqslant q_{on}^{(G)}$.

C. Overlapping of different metastable regions and multiple pattern excitation

The metastable states with respect to the SP formation [dashed region in Fig. 11(a)] and to the DGP and GP formation [dashed region in Fig. 11(b)] partially overlap to one another [Fig. 11(c)]. In overlapped metastable states, depending on initial conditions either free flow or one of SP or else one of GP (DGP) can occur.

In Fig. 12(c), depending on the amplitude of an initial short-time perturbation applied in q_{on} either free flow [Fig. 12(d)], or LSP [Fig. 12(e)], or else GP [Fig. 12(f)] is excited. The stability of these three traffic states is noticeably different. Whereas free flow [Fig. 12(d)] and GP [Fig. 12(f)] remain stable during 120 min in ten different calculated realizations, LSP has spontaneously transformed into GP in six of ten realizations.

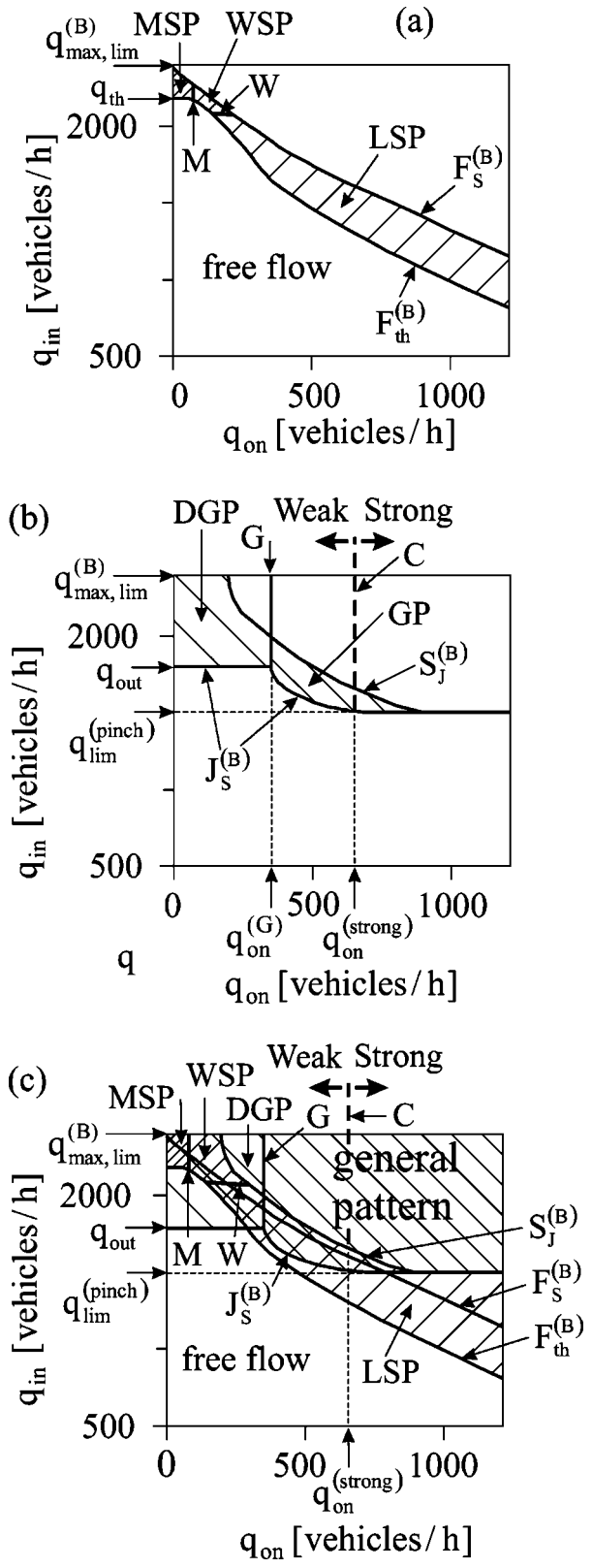

FIG. 11. Metastable regions in the diagram of congested patterns at the on-ramp. (a) Metastable states (hatched regions) where either free flow or SP can occur and exist and also be excited. (b) Metastable states (hatched regions) where GP (right of line G) and DGP (left of line G) can occur and exist and also be excited. (c) The general diagram related to Fig. 5(a) where the regions of the metastable states from (a),(b) are shown.

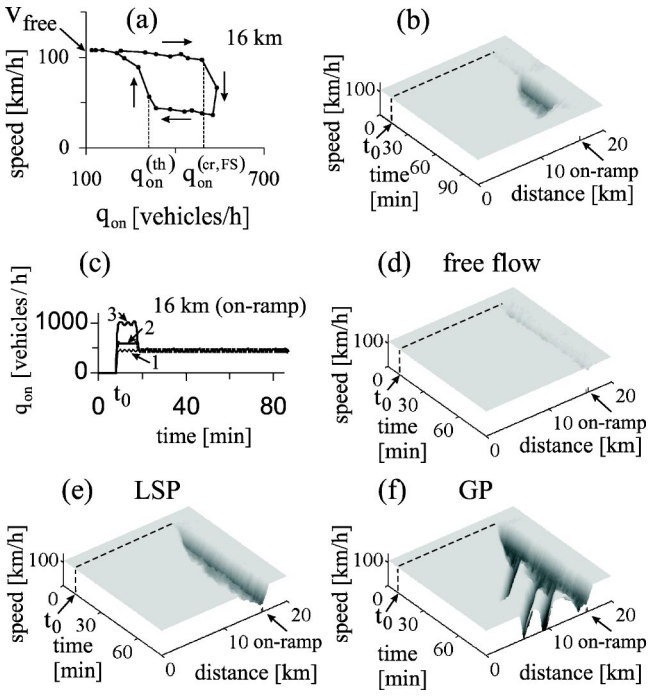

FIG. 12. Hysteresis effect (a), (b) and the multiple metastability at the on-ramp (c)–(f): (a),(b) The 5-min averaged vehicle speed on the main road if q_{on} first increases and then decreases at q_{in} = 1800 vehicles/h (a) and the related SP emergence and dissolving (b). (c) Initial short-time perturbations of q_{on} used for the multiple pattern excitation. (d)–(f) Three different patterns at the same flow rates q_{on} = 450 vehicles/h and q_{in} = 1756 vehicles/h: free flow (d), LSP (e), and GP (f). In (c) no perturbation of q_{on} (line 1) is applied for the free flow occurrence (d); for the excitation of LSP (line 2) and of GP (line 3) q_{on} is increased for 10 min at 600 vehicles/h for LSP (e) and at 1000 vehicles/h for GP (f). The on-ramp parameters [Fig. 1(c)] are: L_m = 500 m, L_r = 300 m. The on-ramp inflow is switched on at $t = t_0 = 8$ min.

VII. STRONG CONGESTION AT MERGE BOTTLENECKS

At the on-ramp when traffic demand (q_{in} and/or q_{on}) is changed weak congestion in GP can transform into strong congestion where pinch region characteristics do not depend on traffic demand anymore (Sec. V). It is found that at the merge bottleneck [Sec. II C; Fig. 1(d)] independent of traffic demand q_{in} *only* strong congestion occurs in GP [Fig. 13(a, b)]. To explain this result, first note that for the bottleneck at the on-ramp the main road at $x < x_{on}^{(e)}$ in Fig. 1(c) may be considered as the left lane of the merge bottleneck at $x < x_M$ in Fig. 1(d) and the on-ramp lane as the right lane of the merge bottleneck. At each of the lanes of the merge bottleneck at $x = -L/2$ the flow rate is equal to q_{in} [Fig. 1(d)]. If now at the on-ramp $q_{on} = q_{in}$ then GP with strong congestion condition is formed [Fig. 5(a)].

However, at the merge bottleneck at $x < x_M^{(s)}$ vehicles can change lane between the left and the right lanes [Fig. 1(d)]. In contrast, at the on-ramp at $x < x_{on}$ the main road and the on-ramp lane are separate roads [Fig. 1(c)]. This may explain the following differences between GPs at the merge bottleneck and at the on-ramp.

(i) In contrast to GP under strong congestion at the on-

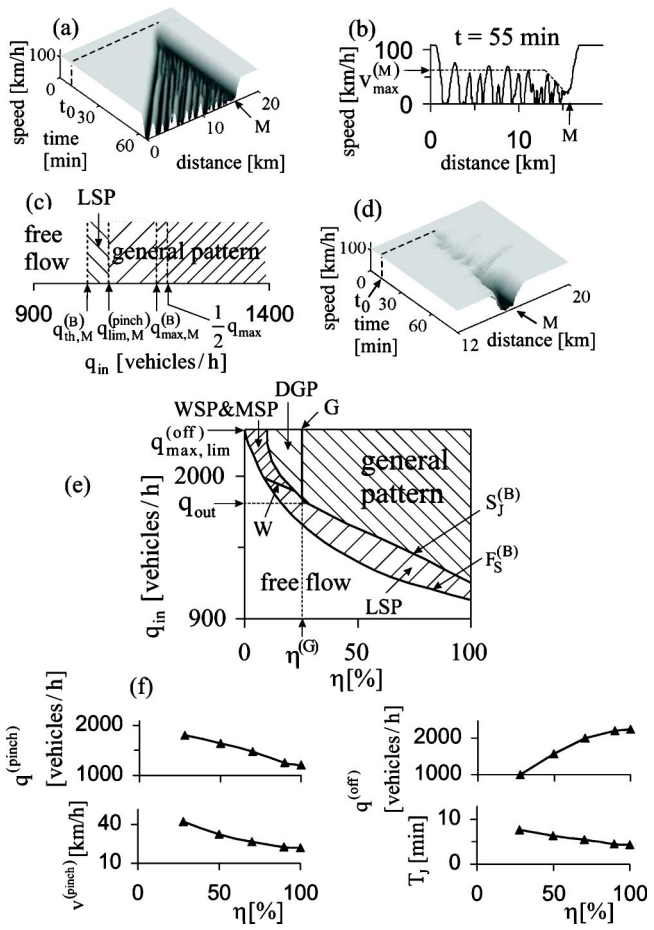

FIG. 13. Features of congested patterns at the merge bottleneck (a)–(d) and at the bottleneck due to the off-ramp (e),(f): (a),(b) Speed distributions in GP. (c) The diagram of patterns at the merge bottleneck. (d) The critical perturbation development that leads to the spontaneous GP formation. (e) The diagram of patterns at the off-ramp. (f) Averaged characteristics of GP at the off-ramp as functions of η. The flow rate to the merge bottleneck q_{in} is (a),(b) 1946 and (d) 1180 vehicles/h. In (a),(d) the vehicle motion on the right lane upstream of the merge bottleneck is switched on at $t=t_0=8$ min. In (a),(b),(d) the letter *"M"* marks the location of the merge bottleneck.

ramp, the average speed between moving jams in GP at the merge bottleneck has a *saturation* to a limit (maximum) speed $v_{\text{max}}^{(\text{M})}<v_{\text{free}}$.

The actual speed $v^{(\text{M})}$ between the moving jams shows only a small amplitude oscillation around $v_{\text{max}}^{(\text{M})}$ [Fig. 13(b)].

(ii) The limit flow rate per highway lane at the merge bottleneck $q_{\text{lim M}}^{(\text{pinch})}$ is considerably lower than $q_{\text{lim}}^{(\text{pinch})}$ at the on-ramp.

(iii) At the same model parameters the frequency of the narrow moving jam emergence $f_{\text{max}}^{(\text{M})}$ is higher and the mean time distance between the jams, $T_{\text{J}}^{(\text{M})}=1/f_{\text{max}}^{(\text{M})}$, is lower at the merge bottleneck ($T_{\text{J}}^{(\text{M})}\approx 3.5$ min at about 1.5–2 km upstream of the bottleneck) than at the on-ramp ($T_{\text{J}}\approx 5-7$ min). The distances between the downstream fronts of wide moving jams in GP, $T_{\text{J}}^{(\text{wide M})}$, are lower at the merge bottleneck ($T_{\text{J}}^{(\text{wide M})}\approx 7$ min at about 10 km upstream of the bottleneck) than at the on-ramp ($T_{\text{J}}^{(\text{wide})}\approx 10-30$ min). Thus, moving jams in GP at the merge bottleneck often look like narrow moving jams. Nevertheless, after the speed saturation has occurred between the moving jams in GP at the merge bottleneck, the jams downstream fronts propagate with the characteristic velocity v_{g} through other upstream bottlenecks and through other states of traffic flow *keeping* the velocity of the downstream front v_{g} (Sec. X A), i.e., the jams are wide moving jams.

The pattern emergence at the merge bottleneck shows the following pattern diagram [Fig. 13(c)].

(i) If $q_{\text{in}}\geqslant q_{\text{max, M}}^{(\text{B})}$, GP can usually occur spontaneously within a given time interval T_0. At $T_0=60$ min the maximum flow rate $q_{\text{max, M}}^{(\text{B})}\approx 1150$ vehicles/h$<q_{\text{max}}/2$ (the factor 1/2 appears because at the merge bottleneck *two*-lane road transforms into *one*-lane road at $x\geqslant x_{\text{M}}$). There is a random time delay for the GP formation after traffic flow has been switched on. A lot of local perturbations can spontaneously appear, grow, and then decay at the merge bottleneck before the perturbation appears whose growth indeed leads to the GP formation [Fig. 13(d)]. However, in some realizations in the range $q_{\text{max, M}}^{(\text{B})}<q_{\text{in}}<q_{\text{max}}/2$ GP does not occur spontaneously during $T_0\approx 180$ min. Instead, an occurrence and disappearance of high amplitude perturbations at the bottleneck is observed. At $q_{\text{in}}\geqslant q_{\text{max}}/2$ GP occurs within $T_0=30$ min.

(ii) If $q_{\text{th M}}^{(\text{B})}\leqslant q_{\text{in}}<q_{\text{max, M}}^{(\text{B})}$ free flow is metastable: (1) At $q_{\text{lim M}}^{(\text{pinch})}\leqslant q_{\text{in}}<q_{\text{max, M}}^{(\text{B})}$ GP can be excited by a high amplitude short-time external perturbation; (2) at $q_{\text{th M}}^{(\text{B})}\leqslant q_{\text{in}}\leqslant q_{\text{lim M}}^{(\text{pinch})}$ rather than GP only LSP can be excited. At $q_{\text{in}}<q_{\text{th M}}^{(\text{B})}$ free flow is stable.

VIII. WEAK CONGESTION AT OFF-RAMPS

A study of the pattern formation at an off-ramp shows the following.

(i) In contrast to the on-ramp and the merge bottleneck, at an off-ramp [Fig. 1(e)], GP *only* under weak congestion can be formed at all $\eta<100\%$.

(ii) In GP, wide moving jams are often formed at a considerable distance (more than 5 km) upstream from the off-ramp. The downstream front of SP and GP is located at some distance (about 1–1.5 km) upstream from the off-ramp. Besides, the speed synchronization across both highway lanes occurs only at some finite distance upstream of $x=x_{\text{off}}$ [Fig. 1(e)].

(iii) The pattern diagram at the off-ramp in the plane (η,q_{in}) [Fig. 13(e)] qualitatively resembles the diagram at the on-ramp [Fig. 5(a)]: There are different SP (MSP, WSP, and LSP), DGP, and GP which look like the related patterns at the on-ramp. However, there is no saturation of the boundary $S_{\text{J}}^{(\text{B})}$ at higher η. This is linked to that in GP only weak congestion occurs at all $\eta<100\%$. Indeed, all pinch region characteristics in GP as well as the flow rate $q^{(\text{off})}$ (vehicles which really leave the main road to the off-ramp) do depend on η [Fig. 13(f)].

(iv) At the same model parameters the values $q^{(\text{pinch})}$, $v^{(\text{pinch})}$, and T_{J} in GP at the off-ramp at high η can be considerably *lower* [Fig. 13(f)] than $q_{\text{lim}}^{(\text{pinch})}$, $v_{\text{lim}}^{(\text{pinch})}$, and $T_{\text{J,lim}}$,

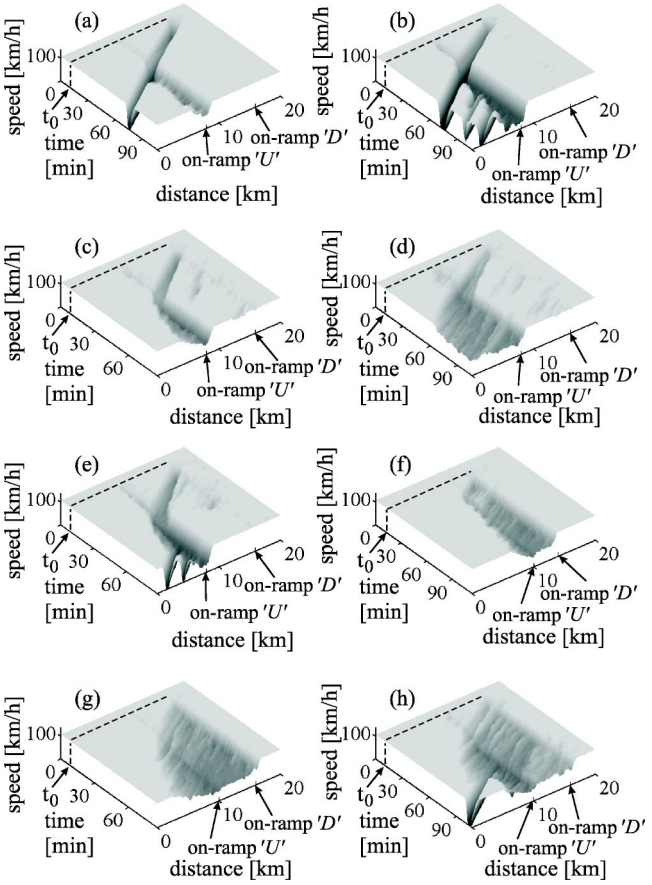

FIG. 14. The induced pattern formation and the catch effect: A wide moving jam induces LSP (a) and GP (b). MSP induces LSP (c), WSP (d), and GP (e). WSP induces LSP (f), WSP (g), and GP (h). The initial flow rates to the on-ramp '*U*' and on the road upstream of the on-ramp '*U*' ($q_{\rm on}^{\rm (up)}$,$q_{\rm in}$) are (a) (320,1800), (b) (450, 1756), (c) (400,1800), (d) (200,2118), (e) (440,1800), (f) (300, 1756), (g) (150,2034), and (h) (240, 2000) vehicles/h. In (a)–(e) the flow rate to the on-ramp '*D*' $q_{\rm on}^{\rm (down)}=60$ vehicles/h. In (f)–(h) $q_{\rm on}^{\rm (down)}$ are 170 (f)–(g) and 180 (h) vehicles/h. In (a)–(h) the coordinates of the on-ramps ($x_{\rm on}^{\rm (up)}$,$x_{\rm on}^{\rm (down)}$) are (a)–(e) (8,16), (f) (10, 14), and (g), (h) (10,16) km. In (a)–(e) the wide moving jam and MSP at the on-ramp '*D*' are induced by short-time perturbations of the flow rate $q_{\rm on}^{\rm (down)}$. $L_{\rm r}=300$ m for both on-ramps, $L_{\rm m}$ is 500 m for the on-ramp '*U*', and 300 m for the on-ramp '*D*'. The on-ramp inflows are switched on at $t=t_0=8$ min.

respectively, in GP under strong congestion at the on-ramp [Figs. 9(a, b, g)]. The off-ramp forces vehicles at higher η to much stronger traffic compression.

(v) At $\eta=100\%$ the off-ramp transforms into the merge bottleneck. We find qualitatively the same GP features at $\eta=100\%$ as for the merge bottleneck.

IX. CATCH EFFECT AND INDUCED CONGESTED PATTERN FORMATION

When two spatially separated bottlenecks (the ''upstream'' bottleneck and the ''downstream'' bottleneck) exist on a highway, then a congested pattern which has earlier occurred at the downstream bottleneck can induce another congested pattern at the upstream bottleneck where free flow is realized before. When both bottlenecks are on-ramps (e.g., Fig. 14), the upstream bottleneck will be designated as the on-ramp '*U*' with the coordinate $x_{\rm on}=x_{\rm on}^{\rm (up)}$ and the flow rate $q_{\rm on}=q_{\rm on}^{\rm (up)}$ and the downstream bottleneck will be designated as the on-ramp '*D*' with $x_{\rm on}=x_{\rm on}^{\rm (down)}$ and $q_{\rm on}=q_{\rm on}^{\rm (down)}$. The following effects have been found.

(i) A congested pattern which induces another congested pattern at the upstream bottleneck is a wide moving jam. In this case, *independent* of the pattern type which is induced at the upstream bottleneck, the wide moving jam propagates further upstream *keeping* the downstream jam front velocity $v_{\rm g}$. In Figs. 14(a,b), a wide moving jam occurs in GP at the on-ramp '*D*' whereas free flow is first at the on-ramp '*U*'. Then, dependent on $q_{\rm on}^{\rm (up)}$, the wide moving jam propagation through the on-ramp '*U*' causes either the LSP formation [Fig. 14(a)] or the GP formation [Fig. 14(b)]. The jam leaves the emerging pattern keeping the velocity $v_{\rm g}$ (for a comparison see empirical Fig. 2(b) in Ref. [17]). After the wide moving jam is far away upstream of the on-ramp '*U*' the flow rate $q_{\rm out}$ plays the role of $q_{\rm in}$: $q_{\rm in}=q_{\rm out}$. In this case, at the chosen model parameters there is no WSP and no MSP [Fig. 5(a)], i.e., either LSP or GP can only occur at the on-ramp '*U*'.

(ii) A congested pattern which induces another congested pattern at the upstream bottleneck is SP. In contrast to case (i), *independent* of the pattern type which is induced at the upstream bottleneck, the initial SP (or the SP downstream front) is usually caught at the upstream bottleneck rather than SP propagating further upstream. Following Ref. [17] this effect is called ''the catch effect.'' In Fig. 14(c), MSP which has occurred at the on-ramp '*D*' first propagates upstream. At the on-ramp '*U*' the initial MSP causes the $F\rightarrow S$ transition. MSP does *not* propagate through the on-ramp '*U*' further: The initial MSP is caught by the on-ramp '*U*' and instead LSP occurs at the on-ramp '*U*' [see empirical Fig. 6(a) in Ref. [17]]. Other two examples of the catch effect are shown in Fig. 14(d, e): The initial MSP is caught at the on-ramp '*U*' in both cases. Instead of this MSP either WSP (d) or GP (e) are induced at the on-ramp '*U*'.

If at the on-ramp '*D*' a WSP occurs [Figs. 14(f,g,h)] then due to the catch of WSP at the on-ramp '*U*' either LSP [Fig. 14(f)] or other WSP [Fig. 14(g)] or else one of GP [in Fig. 14(h) a DGP is shown] can be induced. Note that the induced WSP [Fig. 14(g)] possesses different characteristics than the initial WSP. In particular, the absolute value of the negative velocity of the upstream front of the induced WSP is considerably lower than that in the initial WSP.

(iii) The same effects and nonlinear features have been found out if the downstream bottleneck is the off-ramp rather than the on-ramp '*D*'.

X. COMPLEX CONGESTED PATTERNS AND PATTERN INTERACTION

A. "Foreign" wide moving jams

Wide moving jams which have first emerged at the downstream bottleneck and then propagate through another con-

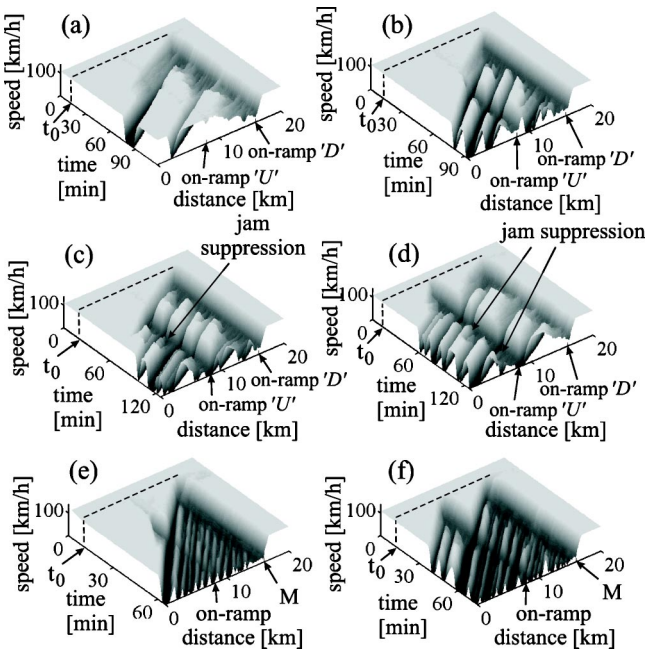

FIG. 15. Foreign wide moving jam propagation. In (a)–(d) $(q_{\rm on}^{\rm (up)}, q_{\rm on}^{\rm (down)}, q_{\rm in})$ are (a) (150,320,2034), (b) (420,500,1675), (c) (600,500,1500), (d) (600,360,1800) vehicles/h. In (e),(f) $(q_{\rm on}, q_{\rm in})$ are: (e) (1200,1800), (f) (2200,2000) vehicles/h. For the merge bottleneck (marked by the letter ''*M*'') in (e),(f) $L_{\rm c}=1000$ m. In (a)–(d) $(x_{\rm on}^{\rm (up)}, x_{\rm on}^{\rm (down)})$ are (8,16) km. In (e),(f) the coordinates of the on-ramp and the merge bottleneck $(x_{\rm on}, x_{\rm M})$ are (8,16) km. The on-ramp inflows in (a)–(f) and the vehicle motion on the right lane upstream of the merge bottleneck in (e),(f) are switched on at $t=t_0=8$ min.

gested pattern which has earlier occurred at the upstream bottleneck are called ''foreign'' wide moving jams [17,26]. Foreign wide moving jam propagation can be shown both in the fundamental diagram approach [11] and within three-phase traffic theory [28]. Apparently the foreign jam propagation as well as the free flow metastability with respect to the $F\rightarrow J$ transition first predicted in [4] are the *only* two effects which can be shown within the fundamental diagram approach according to empirical investigations (see Ref. [27]).

It is found out that the ''foreign'' moving jam propagation can lead to the following nonlinear effects of the pattern interaction and transformation.

(i) In Fig. 15(a), first the upstream front of synchronized flow in GP which emerges at the on-ramp '*D*' is caught at the on-ramp '*U*' and induces WSP as in Fig. 14(g). However, after the first foreign wide moving jam of the GP has later propagated through the on-ramp '*U*', the WSP at the on-ramp '*U*' is suppressed and free flow appears at the on-ramp '*U*' [Fig. 15(a)]. These effects are linked to a decrease in the flow rate upstream of the on-ramp '*U*' due to the foreign moving jam propagation. The initial flow rate $q_{\rm in}=2034$ vehicles/h upstream the on-ramp '*U*' (at which the WSP can exist) is decreased up to the flow rate $q_{\rm out}=1810$ vehicles/h in the foreign moving jam outflow. Another effect is shown in Fig. 15(b). When foreign wide moving jams propagate through LSP at the on-ramp '*U*' the jams change the width and the speed inside LSP only.

(ii) When $q_{\rm on}^{\rm (up)}$ and the flow rate $q_{\rm in}$ upstream of the on-ramp '*U*' are related to the GP nucleation (Sec. VI B) the induced emergence of GP indeed occurs after the first foreign wide moving jam from another GP at the on-ramp '*D*' propagates through the on-ramp '*U*'. However, because in this case $q_{\rm in}<q_{\rm out}$ this foreign jam dissolves very quickly [Fig. 15(c)]. After GP upstream of the on-ramp '*U*' has been formed, other foreign wide moving jams may, nevertheless, propagate further upstream before they dissolve. This is the similar effect as in GP at $q_{\rm in}<q_{\rm out}$ in Fig. 5(g) (Sec. V B).

(iii) If GPs are formed at both on-ramps then narrow moving jams emerging in the pinch region of the upstream GP at the on-ramp '*U*' can be suppressed by foreign wide moving jams from the GP occurring at the on-ramp '*D*' [Figs. 15(c, d)]. Each of the foreign jams suppresses the growth of those narrow jams which are close enough to the downstream front of the foreign jam. Only narrow moving jams which are far enough from the foreign jam may grow and transform into a wide moving jam. Finally a united sequence of wide moving jams occurs upstream of the on-ramp '*U*'. This sequence consists of former foreign wide moving jams and new wide moving jams formed in the GP at the on-ramp '*U*'.

(iv) In Figs. 15(e,f), the downstream bottleneck is the merge bottleneck and the upstream bottleneck is the on-ramp. The moving jams which have emerged in GP at the merge bottleneck propagate through LSP which has spontaneously occurred at the on-ramp [Fig. 15(e)]. The foreign moving jams suppress LSP and fully determine the traffic dynamics upstream of the on-ramp. The jams propagate through the on-ramp keeping the downstream jam velocity $v_{\rm g}$. Thus, the moving jams in GP at the merge bottleneck are indeed wide moving jams (Sec. VII).

(v) In Fig. 15(f), foreign wide moving jams from GP at the merge bottleneck propagate through another GP at the on-ramp. These foreign wide moving jams suppress the growth of all narrow moving jams which would emerge in the pinch region of GP at the on-ramp if there were no foreign wide moving jams. The foreign jams from GP at the merge bottleneck fully determine the traffic dynamics at the on-ramp.

B. Expanded congested patterns

Synchronized flow which has earlier occurred at the downstream bottleneck can first reach the upstream bottleneck and then propagate upstream of the upstream bottleneck. Such a congested pattern is called the ''expanded'' congested pattern (EP) [17]. The downstream front of synchronized flow is fixed at the downstream bottleneck and the upstream front of the synchronized flow is upstream of the upstream bottleneck: Synchronized flow in EP covers both bottlenecks. The following variety of EP features has been found out.

(i) EP can consist of any spatial-temporal combinations between WSPs, LSPs, GPs, DGPs, and foreign wide moving jams. There are nonlinear interactions between these patterns in EP. This interaction can change their features in comparison with patterns features at isolated bottlenecks (Secs.

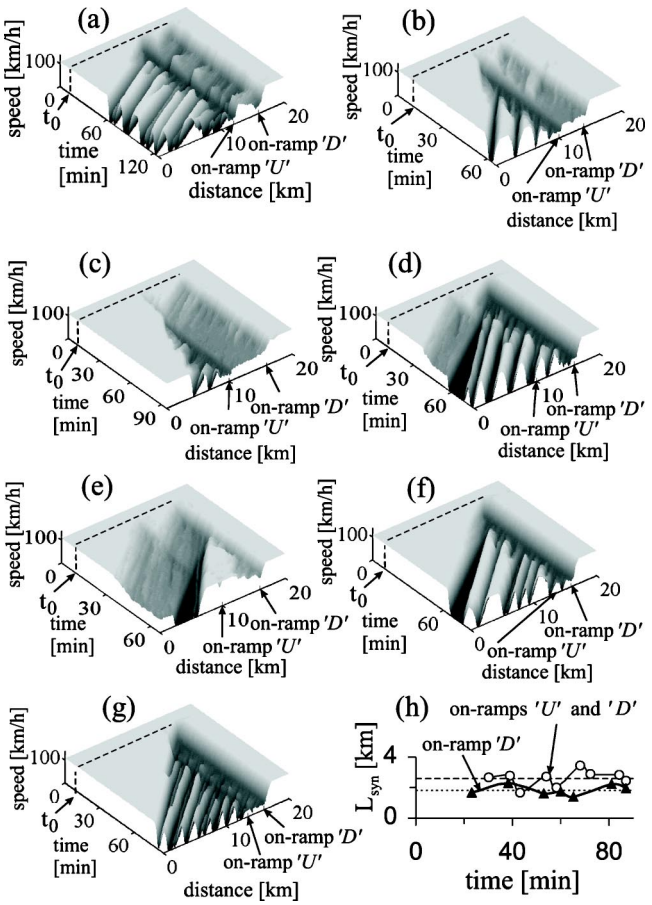

FIG. 16. Expanded congested patterns (EP). (a) GP with two separated pinch regions. (b),(c) EPs which consist of GP at the on-ramp '*U*' and WSP at the on-ramp '*D*'. (d),(e) Dissolving of EP as a result of a suppression of WSP at the on-ramp '*U*' due to the upstream propagation of GP (d) or DGP (e) which has occurred at the on-ramp '*D*'. (f) GP with the pinch region at the on-ramp '*D*'. (g) EP with the pinch region covering both on-ramps. (h) $L_{\mathrm{syn}}(t)$ for GP in (f) (triangles) and for EP in (g) (circles); the mean value $L_{\mathrm{syn}}^{(\mathrm{mean})}$ is shown by dotted line for (f) and by dashed line for (g). $(q_{\mathrm{on}}^{(\mathrm{up})}, q_{\mathrm{on}}^{(\mathrm{down})}, q_{\mathrm{in}})$ are (a) (900,320,1895), (b) (500,170, 1895), (c) (480,190,1745), (d) (195,1000,2285), (e) (175,240, 2285), (f) (0,1000,2000), and (g) (400,1000,1714) vehicles/h. $(x_{\mathrm{on}}^{(\mathrm{up})}, x_{\mathrm{on}}^{(\mathrm{down})})$ are (a) (12,16), (b) (10,14), (c)–(e) (10,16), and (f), (g) (13.5,16) km.The on-ramp inflows are switched on at $t=t_0=8$ min.

V–VIII). In particular, EP can consist of two interacting GPs which are formed upstream of each of two bottlenecks. In Fig. 16(a), two separated pinch regions occur in EP. Wide moving jams which emerge in the GP at the on-ramp '*D*' are foreign wide moving jams for the GP at the on-ramp '*U*'. Synchronized flow covers both on-ramps because the distance between them $L_1=4$ km is not too large. In contrast, if the on-ramps are far enough from one another, EP is not formed. Instead, two different GPs occur so that free flow exists between wide moving jams downstream of the on-ramp '*U*' [Figs. 15(c,d) where $L_1=8$ km].

(ii) EP can consist of WSP at the on-ramp '*D*' and different patterns at the on-ramp '*U*': WSP [Fig. 14(g)], or LSP [Fig. 14(f)], or DGP [Fig. 14(h)], or else GP [Figs. 16(b,c)]. The WSP can influence the patterns upstream because of two effects: (1) WSP can cause the pattern emergence at the on-ramp '*U*' (the catch effect) [Figs. 14(f)–14(h), and 16(c)], and (2) the synchronized flow in WSP can restrict the flow rate and the speed in the outflow of congested patterns at the on-ramp '*U*'. The latter pattern [either WSP, or LSP, or DGP, or else GP, Figs. 14(g,f,h) and 16(b,c)) influences the WSP at the on-ramp '*D*', because the inflow to the WSP is the outflow from the upstream pattern. Note that if the GP under strong congestion occurs at the on-ramp '*U*' [Fig. 16(c)], parameters of the GP only very slightly depend on the existence of WSP. However, the features of the WSP can be changed considerably after the GP has appeared. In contrast with the case of WSP at an isolated on-ramp the average speed in the WSP becomes a stronger falling function of the distance in the upstream direction.

(iii) EP can also occur due to an interaction of WSP at the on-ramp '*U*' with either GP [Fig. 16(d)] or DGP [Fig. 16(e)] at the on-ramp '*D*'. In the first case, rather than EP a GP covering both on-ramps occurs later [Fig. 16(d)]: The initial WSP at the on-ramp '*U*' disappears. Thus, similarly to the GP in Fig. 15(e), the GP in Fig. 16(d) totally determines the traffic dynamics near the on-ramp '*U*'. In the second case [Fig. 16(e)], after the wide moving jam from DGP has reached the on-ramp '*U*', EP disappears, exactly either LSP or free flow occurs at each of the on-ramps. Sometimes, however, EP like that shown in Fig. 14(f) can occur.

(iv) It is found that GP under strong congestion makes the major influence on EP which occurs as a result of an interaction of two patterns at the downstream and upstream bottlenecks. Let us make the following *definition* of *the degree* of strong congestion. The lower the limit flow rate (per highway lane) in GP is, the higher is the degree of strong congestion. It is found that the more the degree of strong congestion in GP is, the more is the influence of this GP on the EP characteristics. An example is shown in Fig. 15(f) where GP at the merge bottleneck has a much more influence on the EP characteristics than GP at the on-ramp.

(v) If the distance between two bottlenecks (due to the on-ramp '*U*' and the on-ramp '*D*') is compatible with the length of the pinch region in GP at the isolated on-ramp (in Fig. 16(f) where $q_{\mathrm{on}}^{(\mathrm{up})}=0$, $L_{\mathrm{syn}}^{(\mathrm{mean})}\approx 1.9$ km) then at $q_{\mathrm{on}}^{(\mathrm{up})}>0$ a GP occurs in which the pinch region covers both on-ramps, i.e., EP occurs [Fig. 16(g)]. The length of the united pinch region in the EP is higher than the one for GP at the isolated on-ramp: $L_{\mathrm{syn}}^{(\mathrm{mean})}\approx 2.7$ km [Figs. 16(h,g)].

XI. INTENSIFICATION OF DOWNSTREAM CONGESTION DUE TO UPSTREAM CONGESTION

It could be expected that the upstream congestion (the congested pattern formation at an upstream bottleneck) should tend to a reduction in the downstream congestion. However, it is found out that due to a complex nonlinear interaction between congested patterns occurring at *different spatially separated* bottlenecks rather than a *reduction* in the downstream congestion an *intensification* of the downstream congestion can occur due to the upstream congestion. The initial flow rate upstream of the on-ramp '*D*' $q_{\mathrm{in}}^{(\mathrm{down})}=q_{\mathrm{in}}+q_{\mathrm{on}}^{(\mathrm{up})}$ in each of the left parts of Figs. 17 is equal to $q_{\mathrm{in}}^{(\mathrm{down})}$

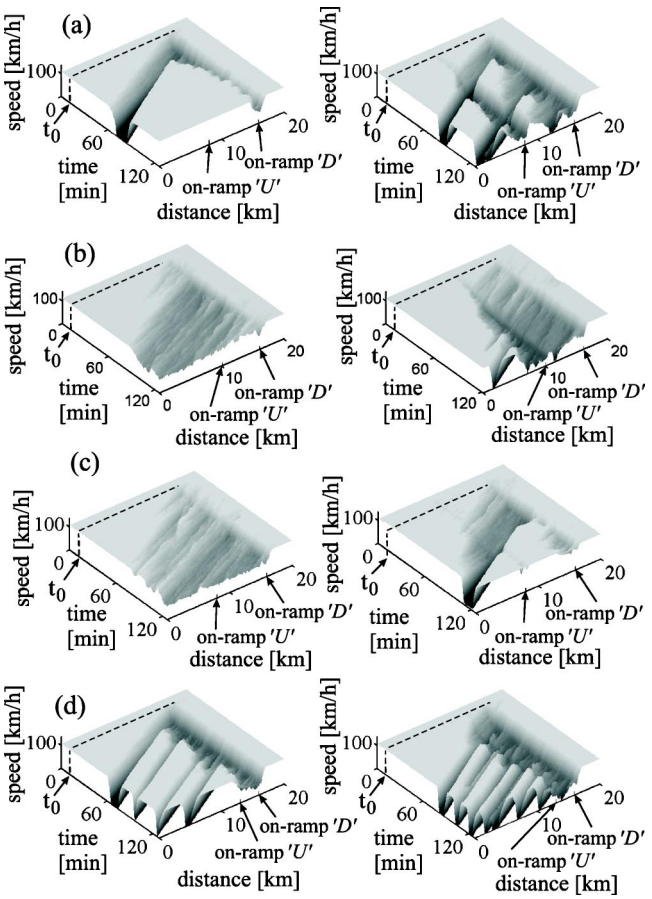

FIG. 17. Intensification of the upstream congestion due to the downstream congestion: (a) DGP (left) transforms into GP (right). (b) In WSP (left) the average speed decreases and a traffic jam emerges (right). (c) WSP (left) transforms into DGP (right). (d) In GP (left) the frequency of the jam emergence increases (right). $t_0=8$ min. In (a)–(d) $(q_{\text{on}}^{(\text{up})}, q_{\text{on}}^{(\text{down})}, q_{\text{in}})$ are (a, left) (0,320,2200), (a, right) (400,320,1800), (b, left) (0,180,2200), (b, right) (420, 180,1780), (c, left) (0,200,2190), (c, right) (155,200,2035), (d, left) (0,365,2200), and (d, right) (300,365,1900) vehicles/h. $(x_{\text{on}}^{(\text{up})}, x_{\text{on}}^{(\text{down})})$ are (a), (c) (8,16), (b) (10,16), and (d) (13,16) km. $L_{\text{r}}=300$ m for both on-ramps.

in each of the related right parts of Figs. 17. However, in Figs. 17, left, $q_{\text{on}}^{(\text{up})}=0$ whereas in Fig. 17, right, $q_{\text{on}}^{(\text{up})}>0$. Dependent on $q_{\text{on}}^{(\text{up})}$, the following phenomena of the intensification of the initial downstream congestion are observed.

(i) The initial DGP at the on-ramp '*D*', where *only one* wide moving jam emerges [Fig. 17(a), left], transforms into GP at this bottleneck where a *noninterrupted sequence* of wide moving jams emerges [Fig. 17(a), right]. To explain this, note that after the wide moving jam in DGP [Fig. 17(a), left] has emerged, the current flow rate upstream of the on-ramp '*D*', $q_{\text{in}}^{(\text{down})}$, has dropped up to q_{out}: $q_{\text{in}}^{(\text{down})}=q_{\text{out}}=1810$ vehicles/h. In contrast, in Fig. 17(a), right, LSP is induced at the on-ramp '*U*' after a wide moving jam from DGP (the downstream congestion) has reached the on-ramp '*U*'. Due to the flow rate to the on-ramp '*U*', $q_{\text{on}}^{(\text{up})}$, the value $q_{\text{in}}^{(\text{down})}$ increases up to $q_{\text{in}}^{(\text{down})}=2150$ vehicles/h. This leads to the further wide moving jam emergence at the on-ramp '*D*', i.e., to the GP formation at the on-ramp '*D*' in Fig. 17(a), right. The mean period of the moving jam emergence in GP is determined by the time of the wide moving jam propagation to the on-ramp '*U*'. Before a wide moving jam has passed the on-ramp '*U*' $q_{\text{in}}^{(\text{down})}=q_{\text{out}}$ and no moving jam can emerge at given $q_{\text{on}}^{(\text{down})}$ at the on-ramp '*D*'.

(ii) The onset of the upstream congestion can lead to a decrease in the average speed inside initial WSP [Figs. 17(b, c), left] at the downstream bottleneck. This intensification of the downstream congestion occurs sometimes with a moving jam emergence in the initial WSP [Figs. 17(b,c), right], i.e., it is accompanied by a transformation of this WSP into a DGP [Fig. 17(c), right] or a GP. The speed in GP at the on-ramp '*U*' is low, which may lead to a decrease in the speed and to the jam emergence in the initial WSP at the on-ramp '*D*'.

(iii) The onset of the upstream congestion can lead to an intensification of the downstream congestion which results in an increase in the mean frequency of the wide moving jam emergence f_{wide} in an initial GP at the downstream bottleneck [Fig. 17(d)]. In Fig. 17(d), right, after a wide moving jam from GP which is under weak congestion at the on-ramp '*D*' has passed the on-ramp '*U*' the current flow rate upstream of the on-ramp '*D*' $q_{\text{in}}^{(\text{down})}$ increases due to $q_{\text{on}}^{(\text{up})}>0$. This leads to the increase in f_{wide}.

The intensification of the downstream congestion due to the upstream congestion has been observed *only* when at the downstream bottleneck weak congestion occurs. This is related to results [17] where diverse transformations between different congested patterns occur at on- and off-ramps under weak congestion.

XII. DISCUSSION

The microscopic traffic flow theory presented in this paper which is based on three-phase traffic theory [23–25] may explain the main empirical features of spatial-temporal traffic patterns at highway bottlenecks [17,25].

(i) Neither a moving jam nor a sequence of moving jams spontaneously emerge in *free flow* at a highway bottleneck if in the vicinity of the bottleneck the vehicle density slowly gradually increases from a low density to higher densities. Rather than the moving jams SP first occurs at the bottleneck. There can be three types of SP: WSP, MSP, and LSP. Alternation of free and synchronized flow can occur in SP. Wide moving jams can spontaneously occur only in synchronized flow. This leads to the GP formation.

(ii) The cases of ''weak congestion'' and ''strong congestion'' should be distinguished. Under strong congestion the average flow rate in GP and characteristics of the moving jam emergence do *not* depend on traffic demand; there is a correlation between outflow from a wide moving jam and the limit flow rate in synchronized flow in GP. When the flow rate to the on-ramp slowly decreases, strong congestion in GP changes to weak one; GP can transform into one of SP. Under weak congestion characteristics of the moving jam emergence in GP depend on traffic demand. Strong congestion often occurs in GP at on-ramps whereas in GP at off-ramps only weak congestion occurs. Under weak congestion

diverse transformations between different congested patterns can occur.

(iii) If two bottlenecks are close to one another, a diverse variety of complex EP, where synchronized flow covers these bottlenecks, can occur. SP emerged at the downstream bottleneck can be caught at the upstream bottleneck (the catch effect). The upstream propagation of either MSP, or of WSP, or else of a foreign wide moving jam emerged at the downstream bottleneck can induce diverse patterns at the upstream bottleneck.

Besides, the theory allows us the following predictions.

(i) At the merge bottleneck [Fig. 1(d)] at lower flow rate LSP can be induced and at higher flow rate GP can occur. GP only under strong congestion can occur at the merge bottleneck.

(ii) At the on-ramp there are diverse regions of metastable states of free flow where at the same flow rates dependent of initial conditions either free flow of one of SPs on else one of GPs can be formed.

(iii) GP with the highest degree of strong congestion makes the major influence on EPs features. Due to the onset of the congestion at the upstream bottleneck an intensification of congestion at the downstream bottleneck can occur.

ACKNOWLEDGMENT

B.S.K. acknowledges funding by BMBF within project DAISY.

[1] G.B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974).

[2] In *Traffic and Granular Flow'99*, Proceedings of the International Workshop on Traffic and Granular Flow, 1999, edited by D. Helbing, H.J. Herrmann, M. Schreckenberg, and D. E. Wolf (Springer, Heidelberg, 2000).

[3] *Traffic and Granular Flow '01* edited by M. Fukui, Y. Sugiyama, M. Schreckenberg, and D. Wolf (Springer, Berlin, 2003).

[4] B.S. Kerner and P. Konhäuser, Phys. Rev. E **50**, 54 (1994).

[5] S. Krauß, P. Wagner, and C. Gawron, Phys. Rev. E **55**, 5597 (1997).

[6] R. Barlovic, L. Santen, A. Schadschneider, and M. Schreckenberg, Eur. Phys. J. B **5**, 793 (1998).

[7] B.S. Kerner, P. Konhäuser, and M. Schilke, Phys. Rev. E **51**, 6243 (1995).

[8] D. Helbing, A. Hennecke, and M. Treiber, Phys. Rev. Lett. **82**, 4360 (1999); M. Treiber, A. Hennecke, and D. Helbing, Phys. Rev. E **62**, 1805 (2000).

[9] H.Y. Lee, H.-W. Lee, and D. Kim, Phys. Rev. E **59**, 5101 (1999); Physica A **281**, 78 (2000); Phys. Rev. E **62**, 4737 (2000).

[10] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, J. Phys. A **33**, L477 (2000).

[11] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, Phys. Rev. E **65**, 015101(R) (2001).

[12] B.S. Kerner and S.L. Klenov, J. Phys. A **35**, L31 (2002).

[13] D. Helbing and M. Treiber, e-print trafficforum/02031301 (2002).

[14] D. Chowdhury, L. Santen, and A. Schadschneider, Phys. Rep. **329**, 199 (2000); K. Nagel, P. Wagner, and R. Woesler, Oper. Res. (to be published).

[15] D. Helbing, Rev. Mod. Phys. **73**, 1067 (2001).

[16] T. Nagatani, Rep. Prog. Phys. **65**, 1331 (2002).

[17] B.S. Kerner, Phys. Rev. E **65**, 046138 (2002).

[18] D.C. Gazis, R. Herman, and R.W. Rothery, Oper. Res. **9**, 545 (1961); R. Herman, E.W. Montroll, R.B. Potts, and R.W. Rothery, *ibid.* **7**, 86 (1959); E. Kometani and T. Sasaki, *ibid.* **7**, 704 (1959); G.F. Newell, *ibid.* **9**, 209 (1961); P.G. Gipps, Transp. Res., Part B: Methodol. **15**, 105 (1981); K. Nagel, and M. Schreckenberg, J. Phys. I **2**, 2221 (1992); M. Bando, K. Hasebe, A. Nakayama, A. Shibata, and Y. Sugiyama, Phys. Rev. E **51**, 1035 (1995); J. Phys. I **5**, 1389 (1995); T. Nagatani and K. Nakanishi, Phys. Rev. E **57**, 6415 (1998); K. Nagel, D.E. Wolf, P. Wagner, and P. Simon, *ibid.* **58**, 1425 (1999); M. Treiber, A. Hennecke, and D. Helbing, *ibid.* **59**, 239 (1999); R. Mahnke and J. Kaupužs, *ibid.* **59**, 117 (1999); E. Tomer, L. Safonov, and S. Havlin, Phys. Rev. Lett. **84**, 382 (2000); E. Tomer, L. Safonov, N. Madar, and S. Havlin, Phys. Rev. E **65**, 065101(R) (2002); R. Kühne, R. Mahnke, I. Lubashevsky, and J. Kaupužs, *ibid.* **65**, 066125 (2002).

[19] A.D. May, *Traffic Flow Fundamental* (Prentice Hall, Inc., New Jersey, 1990).

[20] J. Treiterer, Ohio State University Technical Report PB 246 094, Columbus, Ohio, 1975 (unpublished); M. Koshi, M. Iwasaki, and I. Ohkura, in *Proceedings of 8th International Symposium on Transportation and Traffic Theory*, edited by V.F. Hurdle *et al.* (University of Toronto Press, Toronto, Ontario, 1983) p. 403.

[21] F.L. Hall, V.F. Hurdle, and J.H. Banks, Transp. Res. Rec. **1320**, 91 (1992).

[22] B.S. Kerner and H. Rehborn, Phys. Rev. E **53**, R4275 (1996).

[23] B.S. Kerner, in *Proceedings of the Third Symposium on Highway Capacity and Level of Service*, edited by R. Rysgaard (Road Directorate, Ministry of Transport, Denmark, 1998), Vol. 2, pp. 621–642; in *Transportation and Traffic Theory*, edited by A. Ceder (Elsevier, Amsterdam, 1999), pp. 147–171.

[24] B.S. Kerner, Transp. Res. Rec. **1678**, 160 (1999); Phys. World **12(8)**, 25 (1999); in *Traffic and Granular Flow '99*, Ref. [2], pp. 253–284; J. Phys. A **33**, L221 (2000).

[25] B.S. Kerner, Phys. Rev. Lett. **81**, 3797 (1998).

[26] B.S. Kerner, Netw. Spatial Econ. **1**, 35 (2001).

[27] B.S. Kerner, e-print cond-mat/0211684; and in Ref. [3].

[28] B.S. Kerner, S.L. Klenov, and D.E. Wolf, J. Phys. A **35**, 9971 (2002).

[29] M. Takayasu and H. Takayasu, Fractals **1**, 860 (1993).

[30] S. Krauß, Ph.D. thesis, DRL-Forschungsbericht 98-08.

[31] In Eqs. (1)–(8) the following specific parameters have been used: $\beta=0$, $\epsilon_a(v_n)=0.72-0.55\theta(v_n-15)$, and $\epsilon_b(v_n)=0.95-0.8\ \min(1,v_n/10)$, the other model parameters are the same

as in Sec. II C except $\lambda = 0.5$. Model (1)–(8) is a generalization of KKW cellular automata model [28] for the continuous speed.

[32] Near the threshold density ρ_{th} model fluctuations cause the reverse $S \to F$ transition: MSP can dissolve already at $\rho_{in} > \rho_{th}$ when $(\rho_{in} - \rho_{th})/\rho_{th} \ll 1$.

[33] Different 2D states of synchronized flow are related to different critical perturbation amplitudes. It could be expected that there should be also infinite number of different critical branches for the $S \to J$ transition. The critical branches in Fig. 4 are an averaging of these different critical branches.

[34] At other model parameters the opposite condition $q_{th} < q_{out}$ can be fulfilled. In this case, the range of the MSP existence $q_{th} \leqslant q_{in} < q_{max}$ includes the range $q_{th} \leqslant q_{in} \leqslant q_{out}$.

[35] At $q_{out} \leqslant q_{in} < q_{th}$ and $q_{on} = 0$ SP cannot exist. In this case wide moving jams can be excited in free flow due to application of an external local perturbation of very high amplitude. This amplitude should exceed the critical amplitude $\Delta v^{(B)}_{cr,FJ} = v^{(B)}_{free} - v^{(B)}_{cr,FJ}$ [the critical branch $v^{(B)}_{cr,FJ}$ is shown by the curve 3′ in Fig. 4(e)].

[36] Also in Ref. [12] in MSP which is far away of the bottleneck a wide moving jam can emerge. This effect occurs when model parameters are slightly different from those used for simulations (Sec. II C).

[37] An example of ASP is shown in Fig. 17(c).