Self-organized criticality of traffic flow: There is nothing sweet about the sweet spot

Jorge Laval[a,*]

[a]*School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, United States*

Abstract

This paper shows that the kinematic wave model exhibits self-organized criticality when initialized with random initial conditions around the critical density. A direct consequence is that conventional traffic management strategies seeking to maximize the flow may be detrimental as they make the system more unpredictable and more prone to collapse. Other implications for traffic flow in the capacity state are discussed, such as: (i) jam sizes obey a power-law distribution with exponents 1/2, implying that both its mean and variance diverge to infinity, and therefore traditional statistical methods fail for prediction and control, (ii) the tendency to be at the critical state is an intrinsic property of traffic flow driven by our desire to travel at the maximum possible speed, (iii) traffic flow in the critical region is chaotic in that it is highly sensitive to initial conditions, (iv) aggregate measures of performance are proportional to the area under a Brownian excursion, and therefore are given by different scalings of the Airy distribution, (v) traffic in the time-space diagram forms self-affine fractals where the basic unit is a triangle, in the shape of the fundamental diagram, containing 3 traffic states: voids, capacity and jams. This fractal nature of traffic flow calls for analysis methods currently not used in our field.

Introduction

There is mounting empirical evidence suggesting the hypotheses that urban networks exhibit self-organized criticality (SOC) (Zeng et al., 2019, Zhang et al., 2019, Zeng et al., 2020, Bak et al., 1987), a celebrated paradigm in the 90's that has proved very useful in countless areas of science, from earthquakes to stock prices to forest fires to astrophysics (Dănilă et al., 2015, Aschwanden, 2014). Despite being around for 2 or 3 decades now, the theory and control consequences of SOC have not permeated the transportation literature. This is unfortunate because *if* it turns out that urban networks do exhibit SOC, a whole new paradigm emerges for understanding urban congestion, based on the lessons learned from complexity, catastrophe theory and non-equilibrium critical phenomena (Taleb, 2020). It would mean for example that most of our current control methods based on standard calculus and statistics techniques may no longer apply or may need a major overhaul. Notably, this paradigm is in stark contrast to the currently accepted view in our field that an optimal strategy is to meter the network so that it remains at the "sweet spot": near the critical density where flows are maximized (Geroliminis et al., 2012, Hajiahmadi et al., 2013, Haddad et al., 2013, Ramezani et al., 2015, Keyvan-Ekbatani et al., 2015, Guo and Ye, 2016, Ampountolas et al., 2017, Haddad, 2017, Ding et al., 2017, Fu et al., 2017, Kouvelas et al., 2017, Sirmatel and Geroliminis, 2017, Yang et al., 2017, Yildirimoglu et al., 2018, Zhong et al., 2018, Guo and Ban, 2020).

But given the absence of a general theory of SOC to date (Aschwanden, 2014), it is not yet possible to formally prove or disprove this hypothesis. This paper is a step towards such a theory for traffic flow. The defining characteristic of a SOC system is that it is naturally driven to be in a "critical state" where it undergoes a phase transition (Halperin and Hohenberg, 1969) (e.g. from liquid to solid or from free-flow to congestion). At the critical point the characteristic length scale of the system, such as the size S of forest fires or the size of traffic jams as shown here, diverges to infinity,

---

*Corresponding author
Email address:* `jorge.laval@ce.gatech.edu` (Jorge Laval)
[1]790 Atlantic Dr NW, Atlanta, GA, 30313

which renders the system scale-free and thus scale invariant. This implies a power-law distribution for the size, whose tail probability is proportional to $s^{-\alpha}$, i.e.:

$$\Pr\{S > s\} \sim s^{-\alpha}, \tag{1}$$

with the exponent α known as the critical exponent. As opposed to the normal distribution, where it is virtually impossible to find realizations more than 3 standard deviations away from the mean value, power laws are said to have "heavy tails": there is a non-negligible probability of observing extremely large values. As a consequence, the moments of the distribution may diverge to infinity, in which case the central limit theorem and the law of large numbers no longer hold. When $\alpha > 2$, both the mean and variance are finite; when $1 < \alpha \leq 2$, as it is typically the case, the mean is finite and the variance diverges, and for $\alpha \leq 1$ both the mean and variance diverge. The case $\alpha = 1$ is popularly known as Zipf's law (Zipf, 1949), and as it turns out $\alpha \approx 1 \pm 0.2$ for a wide variety of phenomena, including the size of cities (Mori et al., 2020) and companies (Kobayashi et al., 2021).

SOC behavior in traffic flow models has been conjectured since the mid-nineties in the physics community (Nagel and Paczuski, 1995, Paczuski and Nagel, 1996). Let us define a *jam* as a congestion cluster in the time-space (t, x) plane, i.e. a (t, x)-region of connected occupied "cells" representing stopped vehicles. Using ideas from the theory of percolation phase transitions they show that (i) the critical exponent of jam durations is very close to 1/2, strongly suggesting a relationship with the first return time of a one-dimensional random walk, and (ii) the number of vehicles under congestion scales with time as $t^{1/2}$. They conclude that SOC implies that conventional traffic management strategies seeking to maximize the flow may be detrimental as they make the system more unpredictable and hard to control.

Surprisingly, the impact of SOC in our field has been underwhelming (Nagel, 2021, Schadschneider, 2021). Apart from (Nagel et al., 1996, Rieser and Nagel, 2008, Olmos et al., 2018) who present some numerical evidence of SOC on networks, it appears that the subject has gone largely under the radar. One reason might be that the main insight from the early works (Nagel and Paczuski, 1995, Paczuski and Nagel, 1996) was that SOC in single-lane models emerges only in "cruise-control" type traffic flow models, which only allow random perturbations in congested traffic states. This apparent dependence on the rule details might have been interpreted as the absence of a universality class in traffic flow, implying that SOC might not be an intrinsic property of traffic flow, but just of particular models. Fortunately, recently de Gier et al. (2019) showed that the NaSch CA model (Nagel and Schreckenberg, 1992) and variants that include lane changes belong to the Kardar-Parisi-Zhang (KPZ) universality class (of parameter 3/2). This importance result has the potential of establishing SOC for traffic flow on firm grounds, provided that one can show (i) critical behavior at the transition between free-flow and congestion, and (ii) the system has a tendency to be in the critical state. Unlike (i), the tendency to be in the critical state can be simply explained by the fundamental diagram of traffic flow, where the maximum throughput is observed and the critical density.

This paper presents a detailed analysis of (i) above, based on recent results by (Laval, 2021) showing that the structure of congestion clusters in elementary cellular automata (ECA) rule 184 (Wolfram, 1984) exhibits power-law behavior with an exponent $\alpha \approx 1/2$ at the critical density. This fundamental feature of traffic flow, which appears to have been overlooked in the literature so far, has important implications, which are discussed in detail. We show this by focusing on the fundamental unit of analysis: a single-lane uninterrupted road segment, where the only ingredients are the traffic flow model and the initial conditions. A solid understanding of this fundamental system is important to understand its consequences at the network level.

The remainder of the paper is organized as follows. Section 2 below formalizes the connection between traffic models and random walks conjectured in (Nagel and Paczuski, 1995) using the variational approach to traffic flow. This connection induces a power-law distribution for jam durations, as noted in that reference, but also of the time-space area of a jam. Section 3This presents aggregate measures of performance such as the total distance traveled, average flows and speeds, and probability of congestion. Note that these measures of performance are not heavy tailed, and are related to the area under a random walk. A discussion and outlook is provided in the last section.

1. Universality of traffic flow

Possibly the most remarkable property of some physical systems is universality, stating that critical exponents characterizing the phase transition are determined by the dimensionality of the embedding space and by the degrees of freedom of the variables in the system, but not by the details of the rules governing it (Schroeder, 2009). This is

intimately related to dimensional analysis, commonly used in our field, where the celebrated Buckingham-Pi theorem (Buckingham, 1914) states that a physical system may be fully described in terms of dimensionless combinations of variables, and that typically the describing function is independent of the microscopic details of the problem.

As seen in the introduction, recently de Gier et al. (2019) showed that the NaSch CA model (Nagel and Schreckenberg, 1992) and multi-lane variants that include lane changes belong to the Kardar-Parisi-Zhang (KPZ) universality class of parameter 3/2. This parameter represents the exponent that relates the scaling of space and time variables as $t \sim x^{3/2}$. For example, the relaxation time of a road segment is proportional to (road length)$^{3/2}$, and the correlation length is proportional to $t^{2/3}$.

This result is more general than it looks, because the NaSch model is closely related to many other traffic flow models. In particular, it is known to be a stochastic version of Newell's car following (Newell, 2002) in discrete space, which in turn corresponds to the the kinematic wave model of (Lighthill and Whitham, 1955, Richards, 1956) with triangular fundamental diagram. The fundamental diagram can be readily measured in the field (Newell, 2002); e.g., the critical density free-flow speed and the wave speed. The parameter v_{max} in the NaSch model actually represents the ratio of free-flow speed to the wave speed. When $v_{max} = 1$ the deterministic NaSch model is effectively equivalent to ECA-184, which also corresponds to the kinematic wave model with isosceles fundamental diagram.

Notice that the proof of universality in de Gier et al. (2019) was shown analytically for $v_{max} = 1$ but only numerically for $v_{max} > 1$. One can strengthen this latter result by recalling that Laval and Chilukuri (2016) showed that the slopes of the fundamental diagram are irrelevant. In fact, exploiting a skewing symmetry in the kinematic wave model they showed that the numerical values for these parameters are irrelevant when it comes to the calculation of normalized flows and/or delays. This means that, effectively, traffic flow on a single link requires no parameters (after proper normalization) and therefore is independent of the details such as the value of v_{max}.

Given all of the above, the conjecture becomes that possibly all traffic flow models belong to the same universality class. After all, they all obey the conservation law and have some sort of nonlinear fundamental diagram. Second order effects, bounded accelerations, lane changes and other driving considerations become details under this macroscopic view of traffic models.

2. Criticality of traffic flow

In this section we show that at the critical density, traffic on a long single-lane road is fully characterized by a standard a Brownian motion in continuous time-space, or a random walk in discreet time-space. Based on our discussion in the previous section, in the remainder of the paper we use the simplest traffic flow model, i.e. the kinematic wave model with isosceles fundamental diagram with the following parameters:

$$\text{free-flow speed} = \text{wave speed} = 1, \text{jam density} = 1, \qquad (2)$$

so that it is equivalent to ECA-184. These parameters imply that both the capacity and the critical density are 1/2, that distances are measured in units of one vehicle length and that time is measured in units of 1/2 of a critical headway.

2.1. Connection with Brownian motions

Here we exploit the fact that the kinematic wave model can be posed as a Hamilton-Jacoby PDE (Daganzo, 2005, Laval and Leclercq, 2013), where the solution to the initial value problem with data $G(x), 0 \leq x \leq L$, is given by the celebrated Hopfs-Lax formula (Hopf, 1970). For a generic time-space point $P \equiv (t, x)$, this formula gives $N(P)$, the cumulative vehicle number having crossed location x by time t, which in the case of the isosceles fundamental diagram (2) reads:

$$N(P) = \min_{x_U \leq y \leq x_D} f_P(y) \equiv G(y) + (t - x + y)/2, \qquad (3)$$

subject to initial conditions $N(0, x) = G(x)$, and where $x_U = x - t, x_D = x + t$ are the spatial coordinates of points U and D in Fig. 1. Variational formulation (3) is solved by finding the minimum straight-like path from P to the initial boundary at $(0, y)$, where the cost of each path, $f_P(y)$, is the sum of the terminal cost $G(y)$ and the maximum number of vehicles that could possibly cross this path, $(t - x + y)/2$ in our case.

With deterministic initial conditions at the critical density we have

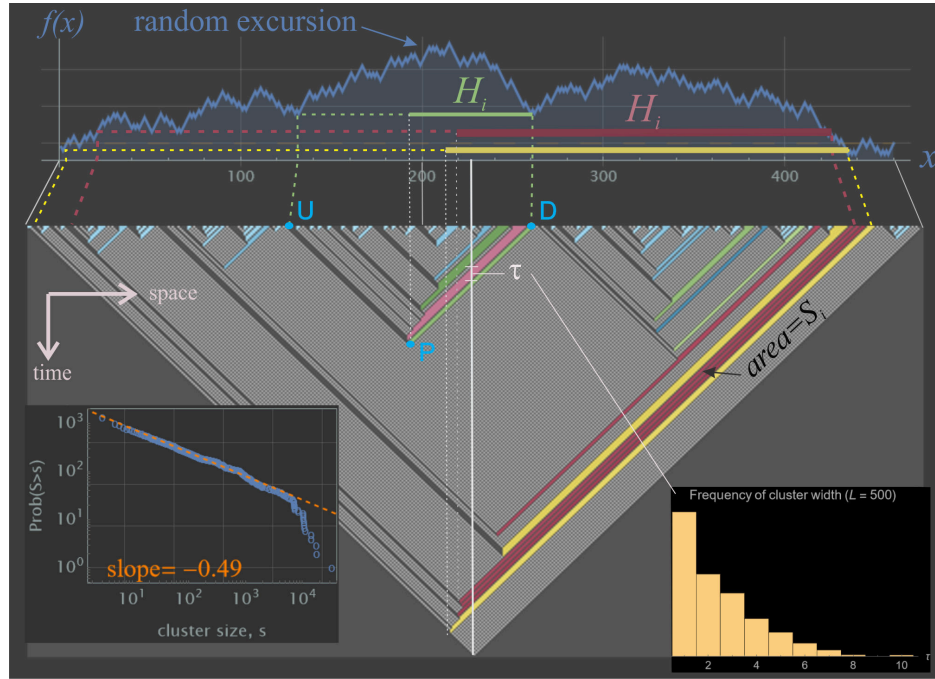$$G(y) = (L - y)/2, \qquad 0 \leq y \leq L, \qquad (4)$$

3

Figure 1: Correspondence between the area under the random walk and the space-time diagram: Sample spatiotemporal realization for ECA rule 184 under critical random initial conditions along with the cost function $f_P(x)$ which is a standard random walk. The colored bars represent the congestion clusters, the dark gray bars the voids and the light gray areas the capacity state.

so that the cost function becomes $f_P(y) = (L - (x - t))/2$, which is independent of y. Therefore, *all* paths from P to the boundary are optimal (since they all have the same cost) and $N_{\text{det}}(P) \equiv (L - (x - t))/2$ is the deterministic solution.

The important point here is that when $G(y)$ is a stochastic process, at the critical density we still have that its expected value $E\{G(y)\} = (L - y)/2$, and that the expected cost is y-independent:

$$E\{f_P(y)\} = E\{G(y)\} + (t - x + y)/2 = (L - (x - t))/2. \tag{5}$$

This y-independence implies that the random process $f_P(y)$ is stationary and with 0 drift. Assuming independent increments of initial conditions leads directly to a Brownian motion process for $f_P(y)$, and will be the working assumption in this paper. Notice that the factor "1/2" in (3), (4) and (5) is the critical density and implies that $f_P(y)$ is normal with mean $(L - (x - t))/2$ and variance $(L - (x - t))/4$.

The independent increments assumption is convenient not only for tractability, but because among all stationary stochastic processes, the Brownian motion is akin to a first-order stochastic approximation. Therefore, the insights revealed by this solution is likely to encode most of the important features of correlated initial conditions (as we have corroborated; see the discussion section). As shown next, the tractability afforded by this assumption means that one can leverage many important results concerning Brownian motions to derive key properties of traffic flow.

2.2. Marginal distribution of the N-surface

From the variational formulation (3) it becomes clear that the distribution of the vehicle number $N(P)$ is given by the minimum of the Brownian motion over the finite interval $x_U < y < x_D$. Conditional on the Brownian motion being zero at the beginning of the interval, $f_P(U) = 0$, this minimum is known to have the half-normal distribution with parameter $\sqrt{\pi/2(x_D - x_U)}$ in our case. Since $x_D - x_U = 2t$ it follows that the probability density of $N(P)$ conditional on $f_P(U) = n_U$ is a shifted half-normal distribution given by:

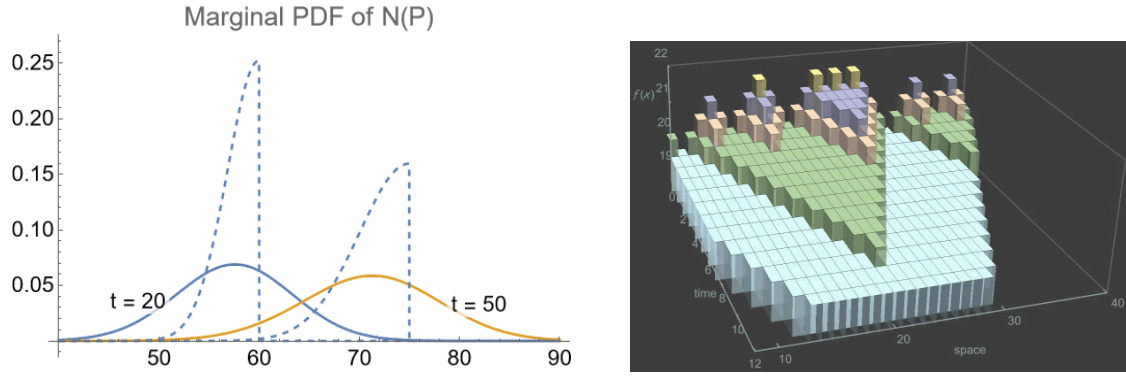$$\phi_{N(P) \mid n_U}(n) = \frac{2}{\sqrt{\pi t}} \exp\left(-(n - n_U)^2/t\right), \qquad n \le n_U \tag{6}$$

4

Figure 2: Left: Marginal distribution for $N(P)$ using $L = 200, x = L/2$ for selected times. For comparison, the half-normal distribution (6) is included as dashed lines. Right:3D representation of the N-surface in oblique coordinates, where the third axis corresponds to deviations around the deterministic mean of the N-surface: $\Delta N(P) = N(P) - N_{\text{det}}(P)$, with $N_{\text{det}}(P) = (L - (x - t))/2$. The random walk can be seen as a slice of the deviation surface at $t = 0$, and the area under the random walk as the projection of the surface into the $(\Delta N, x)$-plane; note that state C is lost in this projection.

with mean $E\{N(P) \mid n_U\} = n_U - \sqrt{t/\pi}$. A natural choice for n_U would be $E\{N(U)\} = (L - (x - t))/2$:

$$E\{N(P) \mid n_U = (L - (x - t))/2\} = (L - (x - t))/2 - \sqrt{t/\pi} \tag{7}$$

and one can see that the stochastic solution is below the deterministic one, i.e. $E\{N(P) \mid n_U\} < N_{\text{det}}(P)$, as expected from Jensen's inequality, but with the discrepancy $\sqrt{t/\pi}$ increasing only as $\sim \sqrt{x_D - x_U}$. The distribution (6) is illustrated by the dashed lines in Fig. 2.

The unconditional marginal distribution for $N(P)$ is obtained by taking expectation of (6) with respect to the distribution of n_U, which is normal with mean $\nu = (L - (x - t))/2$ and variance $(L - (x - t))/4$:

$$\phi_{N(P)}(n) = (\pi(\nu + t))^{-1/2} \exp\left(-\frac{(n - \nu)^2}{\nu + t}\right) \text{erfc}\left((n - \nu)\sqrt{\frac{t}{\nu(\nu + t)}}\right) \tag{8}$$

Unfortunately, analytical expressions for the moments of this distribution are unavailable at this time. Fig. 2(Left) illustrates the shape of this distribution at a fixed location and for different times. It can be seen that the distribution is much flatter than the half-normal distribution (6) (dashed lines in the figure), as a consequence of large variance of n_U.

A realization of the N-surface is shown in Fig. 2(Right), which depicts a 3D representation this surface in oblique coordinates where the third axis corresponds to deviations around the deterministic mean of the N-surface: $\Delta N(P) = N(P) - N_{\text{det}}(P)$. The random walk can be seen as a slice of this surface at $t = 0$, and the area under the random walk as the projection of the surface into the $(\Delta N, x)$-plane; note that state C is lost in this projection.

2.3. Flow distribution

A full description of the N-surface distribution, $\Pr\{N(P_1) < n_1, N(P_2) < n_2, \ldots\}$, has proven difficult to obtain, from which the distribution of flows, $q(P)$, could be derived. However, we have verified that the mean of the conditional (6) and unconditional (8) distributions are numerically very similar, and thus the mean flow can be approximated taking the time-derivative in (7), to obtain:

$$E\{q(t)\} \approx (1 - (t\pi)^{-1/2})/2, \tag{9}$$

where we have used $q(t)$ instead of $q(P)$ to emphasize that the flow is location-independent. Notice that this function starts at 0 at $t = 1/\pi$, increases rapidly to 0.4 by $t = 10$ and flattens out to converge asymptotically to the capacity $1/2$. (As a reference, recall that the critical headway is 2 time units.) That the flow is expected to be low at earlier times is a consequence of the system being farther from equilibrium at earlier times, which translates into an increased activity in jam interactions that produce little flow. As time goes by these interactions resolve to produce flows closer
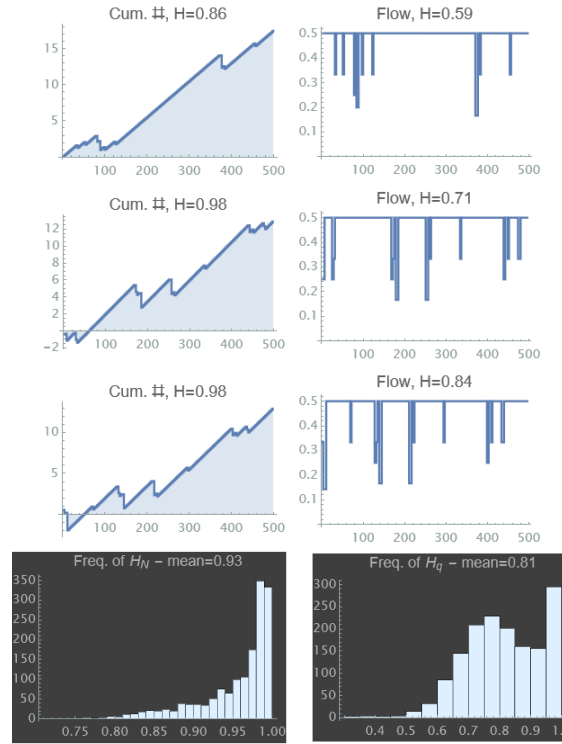
Figure 3: Oblique cumulative count $N(t) = \int_0^t (q(s) - 0.45)ds$ and flow time series at 3 different locations for a typical rollout with $L = 1000$. Each panel shows the Hurst exponent estimated by fitting a fractional Brownian motion process for the cumulative counts and a fractional Gaussian noise process for the flows. The histograms summarizes the results for 2000 rollouts.

to capacity. This phenomenon can be seen in Fig. 1, where it is clear that jam interactions are much more prevalent near the base of the triangle then near the bottom vertex.

To gain additional insights about the distribution of flows, note from Fig. 1 that the flow at each time step can be either 1/2 in the capacity state or 0 in the void and congested states. As shown in the figure, the duration of both the void and congested states at a given location, τ, does not have a heavy-tail distribution and has a mean of approximately 2.7, independent of L. This number is surprisingly small compared to the total number of elementary jams for a particular realization.

Fig. 3 shows the (oblique) cumulative counts and flow time series at 3 different locations for a typical rollout. The roughness of the cumulative curve is apparent as a result of the "spiky" variations in the flow around the capacity value 1/2. As a first approximation to describe the stochastic processes for flows, a fractional Gaussian noise process was fitted for 2000 realizations of flows timeseries. Also, a fractional Brownian motion process was estimated for the cumulative counts. Each panel shows the estimated Hurst exponent and the histograms summarize the results across rollouts. The average Hurst exponent for the flows $H = 0.81$ is in line with the empirical results reported for Australian motorways during morning rush hour and nights (Chand et al., 2017).

These high Hurst exponents indicate a long memory in the time series as a result of long-range spatial correlations. In this case these correlations are due to the spatial interactions brought about by the traffic model, which imposes long periods of constant capacity flow interrupted by bursts of null flow. Notice that the fractal dimension of the flow time series can be approximated by $D = 2 - H = 2 - 0.81 \approx 1.2$.

2.4. Elementary jams

Let us define an *elementary jam* as the collection of occupied cells connecting a point P located on the back of the queue and its corresponding downstream point D on the boundary; see Fig. 1. This is akin to a congested characteristic line connecting P to D but in discrete units. Points P in the back of the queue are characterized by a transition from

6

free-flow to congestion at that point, which means that the cost of minimum paths P-U and P-D both solve (3) and therefore have identical cost. This means that the duration $D - U$ in Fig. 1 is given by the first return time of our random walk $f_p(y)$, which are known to have a power law distribution with exponent 1/2 (see e.g. Schroeder, 2009, Kostinski and Amir, 2016, p.152). From Fig. 1 it is clear that the duration, H, of an elementary jam is $H = (D - U)/2$ and therefore retains the same power-law distribution with exponent 1/2:

$$\Pr\{H_i > h\} \sim h^{-1/2} \tag{10}$$

The consequences of this are far-reaching: both the mean and variance of elementary jam durations (and distance traveled) diverge to infinity, which renders standard statistical tools inadequate. Most notably, the law of large numbers does not apply: taking a sample of jam durations to estimate mean and variance is futile because these estimates will not coverage regardless of the sample size.

2.5. Cluster size distribution

We consider here a road segment divided into L cells, where the initial "density" in each cell $i = 1, 2, \ldots L$ can be either 0 or 1 with probabilities 1-p and p, respectively, independently of other cells (by the independent increments assumption); i.e., a Bernoulli process. Notice that p can also be interpreted as the mean density in the segment.

As mentioned to the introduction, (Laval, 2021) recently showed numerically that ECA 184 exhibits power-law behavior at the critical density, with cluster sizes, S, following a power-law distribution with an exponent close to one half:

$$\Pr\{S > s\} \sim s^{-\alpha}, \quad \alpha \approx 0.5 \pm 0.1 \tag{11}$$

This was shown by evolving ECA-184 for $T = 5,000$ time periods starting with an initial row of $L = 10,000$ cells, each with a probability $p = 1/2$ of being occupied. The time-space area $A = LT/2$ for cluster measurement is of triangular shape as in Fig. 1, where $L = 2T$ to ensure that all points in A are able to reach the boundary. The tail distribution of the site percolation cluster size (Stauffer and Aharony, 2018) were plotted on a log-log plot so that the slope of the linear segment (if any) gives an idea of the exponent of the power law; see inset in Fig. 1 for $p = 1/2$. Piecewise linear regression, with 2 pieces, was used to estimate the power-law exponent in order to account for finite-size effects; see the 2 estimated lines in Fig. 4 and note that the power-law exponent is taken as the slope of the uppermost linear piece, the orange line in the figure.

To study the behavior near the critical density, here we expand the experiment in (Laval, 2021) to include a range of mean initial densities p. Fig. 4 shows the relationship between the estimated power-law exponent α and p in our experiment for $L = 1,000$. The downwards trend is fairly linear around the critical density, as in second-order phase transitions, and can be described approximately as

$$\alpha(p) \approx 1/2 - 4(p - 1/2) \pm 0.1, \qquad 0.45 \le p \le 0.55 \tag{12}$$

The sizable but decreasing interval width of the transition, $\approx \pm 0.1$ at the critical density, is a strong indication that the exponent that best describes a particular realization is highly sensitive to the initial conditions. This makes prediction challenging, especially considering that $L = 1,000$ is a fairly large road segment of around 10 kilometers, and that the interval width increases for smaller segments, e.g. $\approx \pm 0.2$ for $L = 100$ as seen in Fig. 4.

It is important to note that only near the critical density these power-law exponent are meaningful. Fig. 4(Bottom) shows 3 detailed outputs for selected values of p, where one can see that for $p << 1/2$ there is a clear power law relationship but with a very limited range, because in the free-flow jams are very small. For for $p >> 1/2$ the cluster distribution describes a piecewise-linear relationship, as a consequence of several clusters "percolating" to the opposite side of the triangle in the time-space plane.

2.5.1. Intuitive explanation of Eqn. (11)

In terms of elementary jams, a jam can be characterized by a collection of, say n, contiguous elementary jams. The size S of a jam can be seen as the sum of n *correlated* power laws with exponents 1/2:
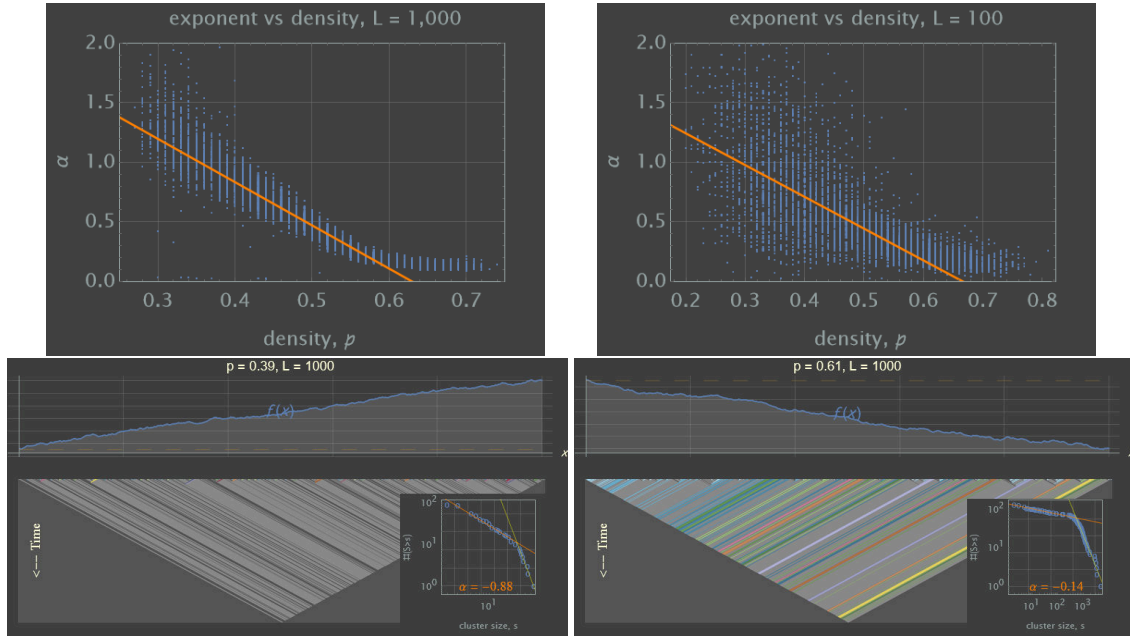
$$S = \sum_{i=1}^{n} H_i. \tag{13}$$

7

Figure 4: Top: Estimated power-law exponent α vs the initial density p for $L = 1,000$ and $L = 100$; orange regression lines $\alpha = 2.476 - 4.022p$ and $\alpha = 1.777 - 2.668p$, respectively, estimated with data in $0.45 \leq p \leq 0.55$. For lower values of the segment length results are similar but with a higher variance. Bottom: Detailed outputs for selected values of $p \neq 1/2$ for $L = 1,000$.

It is well known that for independent H_i's and for large values of S the power law approximation for S is accurate (Zaliapin et al., 2005), and maintains the same exponent, $1/2$ in this case, regardless of the value of n. In light of (11) one concludes, perhaps surprisingly, that the correlations among the H_i's are not a problem that would alter the scaling exponent of cluster sizes.

This can be understood intuitively thanks to (i) the correspondence between the random walk and the jam clusters unveiled in this paper, and (ii) the property of power laws that the conditional distribution $\Pr\{H_i > h | H_i > a\} \sim h^{-1/2}$ is also a power law with the same exponent. To see the implication of (i), take the duration H_i in Fig. 1 that corresponds to points P and D in the figure. One can see that the duration of the elementary jam above it, which is part of the same jam, is smaller than H_i and therefore correlated, but the one below it, which is part of *another* jam, is far greater and uncorrelated given that this particular H_i happens to begin/end at a local minimum of the random walk. We conclude that within a jam the H_i's are correlated, across jams they are not. This means that jam cluster sizes S are uncorrelated.

To see the implication of (ii) above, notice that the number of contiguous elementary jams n can be seen as a random stopping time for the sun in (13), and that, starting with the smallest element in that sum, each consecutive element will be greater than the previous one. Therefore, given property (ii) above, the sought result follows.

3. Aggregate measures of performance

Let $\Psi \equiv \int_A k(t, x)\, dA$ and $\Phi \equiv \int_A q(t, x)\, dA$ be the total time traveled and total distance traveled, respectively, over a (discrete) time-space region A. By letting the symbol A also denote its area, Edie's (Edie, 1965) average density, flow and speed in A, k_A, q_A and v_A, respectively, are given by:

$$k_A \equiv \Psi/A, \quad q_A \equiv \Phi/A, \quad v_A \equiv q_A/k_A \tag{14}$$

These general definitions can be greatly simplified in our case because the spatiotemporal solution to our problem only involves 3 traffic states: capacity state C with (flow, density) = (1/2, 1/2), the empty or void state O with (0,0) and the jam state J with (0, 1). The total area that each state occupies in A is indicated with a subscript, and they satisfy:

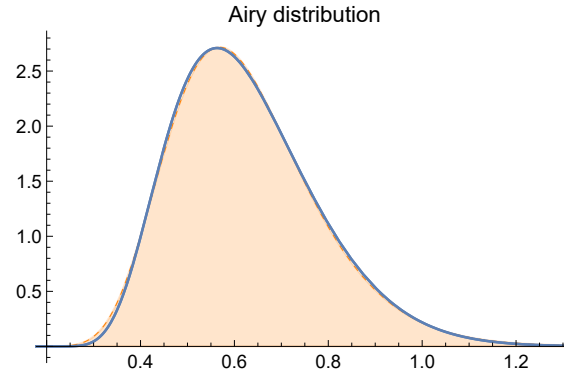$$A = A_0 + A_C + A_J, \quad A_J = A_0 \rightarrow A_C = A - 2A_J \tag{15}$$

8

Figure 5: PDF of the Airy distribution, $g(a), 0 \leq a \leq 1$, in blue. Although it is analytical, there is no closed-form expression for this distribution, only a complicated infinite sum (Majumdar and Comtet, 2005), which is not implemented by default in most libraries/packages (the first 24 terms in this sum were considered for making this figure). The shaded area corresponds to the area under the Inverse Gaussian Distribution with mean parameter 0.625416 and shape parameter 9.82985, which can be seen to provide an accurate (and fast) approximation.

The total travel time $\Psi = \int_{A_C} \frac{1}{2} \cdot dA + \int_{A_J} 1 \cdot dA = A_C/2 + A_J$, and the total distance traveled $\Phi = A_C/2$. Using (15) we have:

$$\Psi = A/2, \quad \Phi = A/2 - A_J \tag{16}$$

and therefore $A_J < A/2$ is bounded. Also,

$$k_A = 1/2, \quad q_A = 1/2 - p_J, \quad v_A = 1 - 2p \tag{17}$$

where $p_J = A_J/A$ correspondence to the probability that a time-space point in A is congested. It can be seen that the total time-space area in congestion, A_J, turns out to be the key measure of performance from which all others can be derived.

3.1. Moments and probability densities

The main result here is that for an area $A = L^2/4$ of triangular shape as in Fig. 1, the area A_J of congested clusters corresponds to half the area under a Brownian *excursion*, $e(x)$:

$$A_J(L) = \frac{1}{2} \int_0^L e(x)dx. \tag{18}$$

A Brownian excursion is a conditioned one dimensional Brownian motion over the time interval $(0, L)$ such that its path starts and ends at the origin and it is constrained to stay positive in between. It is important to note that the area under the excursion is also given by the area between a Brownian bridge and its minimum value in $(0, L)$. This second interpretation is precisely our context: (i) exactly at the critical density the number of empty and unoccupied cells are equal so that in our random walk $f_p(0) = f_p(L)$, which corresponds to a Brownian bridge, and (ii) from Fig. 1 it can be seen that (half) the total area under the random walk and its minimum value is A_J, since:

$$A_J = \sum_{i=1}^{m} H_i \tag{19}$$

where m is the total number of elementary jams in the system. It is important to highlight that this equation remains valid so long as the time-space region of analysis is as in Fig. 1, i.e. an isosceles triangle in the shape of the fundamental diagram, with area $A = L^2/4$. This triangle also corresponds to the domain of dependence of the continuum traffic model, and contains all time-space points that can be affected by perturbations at the boundary.

Unlike (13), the correlations among the H_i's in (19) play a significant role here because the sum across all elementary jams in the system is limited by the size of the system itself, and therefore its distribution will not be a power law, as shown next.

For the unit length road segment the area under a Brownian excursion, $Z = \int_0^1 e(x)dx$ is known to have the Airy distribution (Janson, 2007, Majumdar and Comtet, 2005, Agranov et al., 2020); let $g(a), 0 \leq a \leq 1$ be its PDF. Fig. 5 plots the PDF of the Airy distribution, $g(a), 0 \leq a \leq 1$. Notice that it is not a heavy-tail distribution; (Janson, 2007) shows that its mean and standard deviation are given by $\mathrm{E}\{Z\} = \sqrt{\pi/2}/2 \approx 0.627, \mathrm{SD}\{Z\} = \sqrt{5/12 - \pi/8} \approx 0.155$. Although the Airy distribution is not exactly symmetric, one can approximate the interval where most of the realizations of Z would fall: $Z \in 0.627 \pm 0.155 = (0.472, 0.781)$.

Given the self-similarity properties of Brownian motions, we have that $A_J/L^{3/2}$ has the same distribution as $Z/2$ (Majumdar and Comtet, 2005), and therefore is an invariant quantity in our problem. Most of the realizations of $A_J/L^{3/2}$ would fall in the interval $(0.236, 0.391)$. Similarly, the invariant probability of congestion is $p_J L^{1/2}$ and has the same distribution as $2Z$. Importantly, it also follows that:

$$\mathrm{E}\{A_J\} \sim L^{3/2}, \qquad \mathrm{E}\{p_J\} \sim L^{-1/2}. \tag{20}$$

To correctly interpret these insights, which are not contradictory, recall that the analysis area $A = L^2/4$ changes with the size of the segment. Therefore, doubling the size of the road increases the analysis area by a factor of 4, but the congested area A_J only by a factor of $2^{3/2} \approx 2.8$, implying economies of scale. It is not surprising then, that the probability of congestion actually decreases by a factor of $2^{-1/2} \approx 2.8/4 \approx 0.71$.

To obtain the PDF of A_J one can use the above self-similarity arguments to show that

$$\phi_{A_J}(a) = 2L^{-3/2}g(2aL^{-3/2}) \tag{21}$$

The distribution of all other traffic flow variables will be given by different scalings of the Airy distribution. It follows from (17) that the PDFs for these variables are given by

$$\phi_{p_J}(a) = A\phi_{A_J}(aA) = g(aL^{1/2}/2)L^{1/2}/2 \tag{22a}$$

$$\phi_{q_A}(a) = \phi_{p_J}(1/2 - a) \tag{22b}$$

$$\phi_{v_A}(a) = \phi_{p_J}((1 - a)/2)/2 \tag{22c}$$

which only depend upon $L^{1/2}$. It can be seen from (22a) that the Airy distribution can be interpreted as the likelihood that a particular point in $A = L^2/4$ is congested on a segment of length $L = 4$.

4. Discussion

This paper has presented overwhelming evidence that traffic flow in individual road segments exhibits all the characteristics of critical systems subject to SOC. We have shown that this critical behavior can be traced back to the deep connection with Brownian motions formalized here, a theory well understood by now that you have important implications for traffic flow going forward; we have only scratched the surface here.

This critical behavior added to the tendency of the system to be in the critical state is what makes the case for SOC so compelling. This tendency is a consequence of our driving behavior as we seek to maximize travel speeds–subject to safety–and thereby maximizing the flow out of a jam. The corresponding SOC "sandpile" model is one where vehicles are added one at a time to an initially empty ring road and observing whether or not a slowdown is produced. If not, another vehicle is added until this happens. The first slow down gives an indication of the beginning of the critical region, and congestion clusters of small size are expected. We keep adding vehicles to the system until jams of all sizes are observed, which correspond to "avalanches" in SOC theory.

Starting with random initial conditions at the critical density, we have shown that on long segments congestion clusters propagate for a time (or distance) duration that follows a power law with exponent 1/2. Exponents less than unity are problematic because no moments of the random variable exist, such as the average jam propagation distance (or time) and its variance, which makes prediction virtually impossible as traditional analysis tools are bound to fail. In fact, jams (and voids) of any duration may be expected for no particular reason at the critical density, in the same way earthquakes of any magnitude are expected in some countries. An effective control method is therefore to meter the demand to *prevent* the density from reaching the critical region, and not to maximize throughput as in current practice.
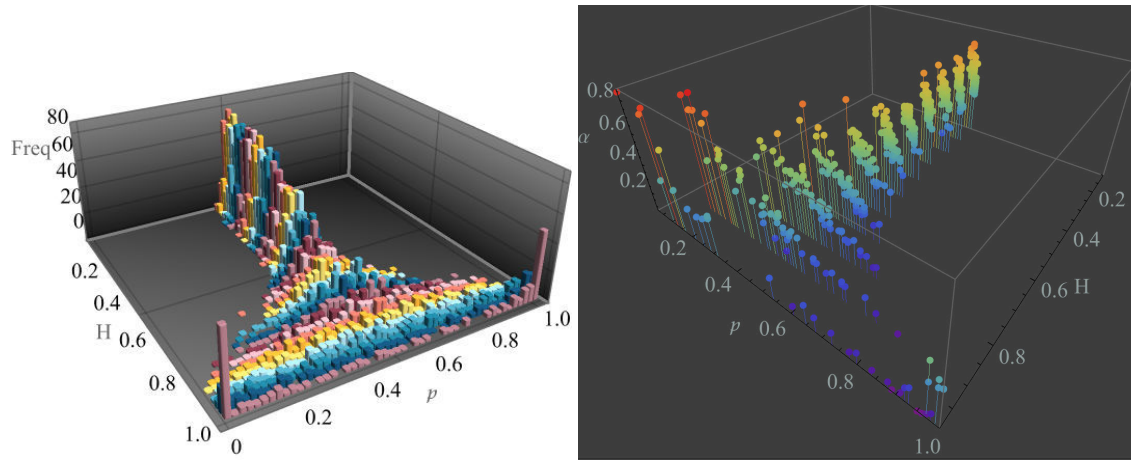
Figure 6: Histogram of occupation probability p vs Hurst exponent H. $L = 200$.

Traffic flow near the critical density is therefore a fractal object, a property that has not been exploited in the current literature. In the (t, x)-plane the basic fractal "generator" unit is a triangle similar in shape to the fundamental diagram, containing 3 traffic states: voids, capacity and jams; see Fig. 1. Within each such triangle all the properties and formulas put forward in this paper apply so long as L is interpreted as the base of the triangle. Similarly, due to the symmetry property of the kinematic wave model, all the results in this paper for the jam state J transfer identically to avoid state O. For example, the distribution of the duration of voids at a given location is identical to the distribution of the duration of jams at the same location; the total area of voids is also half of the area under a Brownian excursion, etc.

This paper has taken this symmetry one step further, by exploiting the connection with the theory of random walks, best illustrated using the traffic flow N-surface as in Fig. 2 (right). Notably, one can see that the area under the random walk can be interpreted as the projection of this surface into the $(\Delta N, x)$-plane, and that state C is lost in this projection. It becomes clear then, that delays at the critical density can be calculated simply by integration of the initial data, without the need to solve the traffic flow model explicitly.

In the case of flow timeseries, we found compelling evidence for a Hurst exponent $H > 0.5$ (so flow fractal dimension $D < 3/2$) indicating long-range correlations for both cumulative counts and flow timeseries, settling the discrepancies in the existing literature between conflicting results (Chand et al., 2017) and (Krause et al., 2017) in favor of the former. It remains to be established, however, that a fractional process approach is justified for traffic flow. From Fig. (3) it is clear that the flow time series is far from a random noise symmetric around its mean value, as required by the fractional process theory. A better description of cumulative accounts and flow time-series can be obtained through fractal geometry, in particular, the relationship $E\{A_J\} \sim L^{3/2}$ unveiled here. This relationship implies a fractal dimension of 3/2 for the cluster areas in the time-space plane, and therefore a fractal dimension of 3/2+1=5/2 for the N-surface, and a fractal dimension of 5/2-1=3/2 for any slice of this surface, such as the cumulative curves or flow time-series.

As mentioned in the introduction, early work (Nagel and Paczuski, 1995, Paczuski and Nagel, 1996) conjectured that SOC in single-lane models emerges only in "cruise-control" type traffic flow models, where random perturbations can only take place in congested traffic states. Here we have found that SOC appears to be an intrinsic property of traffic flow, giving a strong indication that the results presented here should apply in general, as a consequence of our conjecture of universality of the kinematic wave model/ECA-184. These references also pointed out the connection with random walks, based on numerical results that hinted to a power nap exponent close to one half for the duration of a jam. Here we have put this results on firmer grounds thanks to the connection with variational theory presented in section 2.1. A discrepancy found with this earlier work is in the distribution of jem sizes (11), where they found $\alpha = 1/3$ as opposed to one half as shown here. Their result is a direct consequence of their conjecture that the number of vehicles under congestion scales with time as $t^{1/2}$, for which we found no support here.

In this paper we have concentrated on initial conditions of the "white noise" type, which is the simplest paradigm

11

in stochastic analysis as it assumes stationary and independent increments. The next step is to analyze more general initial conditions that allow for spatial correlations, such as fractional Gaussian noise processes with Hurst exponent H, where $H = 1/2$ corresponds to the Brownian white noise. It is important to note that this parameter also determines the occupation probability p. This is shown in Fig. 6 (top), which shows the relationship between the occupation probability p and Hurst exponent H obtained numerically. It can be seen that for $H < 1/2$ the occupation probability follows a narrow normal distribution around a mean of 1/2, and that for $H > 1/2$ this distribution widens as $H \to 1$. This is as expected since long-range dependencies imply that two neighboring cells will tend to have the same value (either occupied or not) as $H \to 1$. The bottom part of the figure shows the power-law exponents and α obtained with these fractional initial conditions and the probation probability p as a function of H. The main conclusion here is that the critical region seems to be wider than in the random walk case, i.e. that apart from the mean initial density p, the spatial correlations in the initial density are also an important factor when it comes to critical behavior. This is currently under investigation by the author.

So, is it time for paradigm change? Most likely yes, at least in the critical region. We have seen that prediction of "local" variables such as individual jam propagation and its effects on flow time-series is virtually impossible in the critical region. The main difficulty here is that traffic flow is "chaotic" in the sense that jam propagation patterns are highly sensitive to initial conditions, i.e. to the individual vehicle arrangement at the beginning of the analysis period. Importantly, trying to predict when a particular road segment will spill back is a lost cause with standard loop-detector technology, but might be achievable with vision technology tracking each individual vehicle. But even then it is possible that random fluctuations will prevent accurate predictions. In the meantime, a robust control strategy at the network level appears to be perimeter control to *prevent* the system for reaching the critical region, as opposed to current practice to keep vehicle accumulation at the "sweet spot" to maximize flows. It turned out that there is nothing sweet about the critical region! That is where catastrophes take place seemingly out of nowhere: e.g., a vehicle not willing to reroute due to a spill back in front, or another one unable to change lanes in time and forced to stop, or a delivery truck suddenly double parking, all might trigger severe congestion in the whole network.

We are therefore in a position now to accept Nagel and Paczuki's 1995 conjecture that traffic management strategies seeking to maximize flow are bound to cause more harm than good. The main reason why flow-maximizing strategies have not proven problematic in the literature might be explained because of the traffic models used, typically some variation of the cell transmission model. While it is true that this model is a numerical solution method for the kinematic wave model, it is unable to capture the nuances of the sensitivity to initial conditions exposed here, which are wiped-out by the mesoscopic averaging over typically long cells and by the considerable numerical error of this model. This argument makes the case for the use of microscopic models (where vehicles do not disappear when unable to take an exit!), but at the expense of increased computation times. Another research direction would be mesoscopic models that capture critical behavior, which appeared to be nonexistent today.

Acknowledgment

References

Agranov, T., Zilber, P., Smith, N.R., Admon, T., Roichman, Y., Meerson, B., 2020. Airy distribution: Experiment, large deviations, and additional statistics. Physical Review Research 2, 013174.

Ampountolas, K., Zheng, N., Geroliminis, N., 2017. Macroscopic modelling and robust control of bi-modal multi-region urban road networks. Transportation Research Part B: Methodological 104, 616–637.

Aschwanden, M.J., 2014. A macroscopic description of a generalized self-organized criticality system: Astrophysical applications. The Astrophysical Journal 782, 54.

Bak, P., Tang, C., Wiesenfeld, K., 1987. Self-organized criticality: an explanation of 1/f noise. Phys. Rev. Lett 59, 381.

Buckingham, E., 1914. On physically similar systems; illustrations of the use of dimensional equations. Physics Reviews , 345–376.

Chand, S., Aouad, G., Dixit, V.V., 2017. Long-range dependence of traffic flow and speed of a motorway: Dynamics and correlation with historical incidents. Transportation Research Record 2616, 49–57.

Daganzo, C.F., 2005. A variational formulation of kinematic wave theory: basic theory and complex boundary conditions. Transportation Research Part B 39, 187–196.

Dănilă, B., Harko, T., Mocanu, G., 2015. Self-organized criticality in a two-dimensional cellular automaton model of a magnetic flux tube with background flow. Monthly Notices of the Royal Astronomical Society 453, 2982–2991.

Ding, H., Guo, F., Zheng, X., Zhang, W., 2017. Traffic guidance–perimeter control coupled method for the congestion in a macro network. Transportation Research Part C: Emerging Technologies 81, 300–316.

Edie, L.C., 1965. Discussion of traffic stream measurements and definitions, in: 2nd Int. Symp. on Transportation and Traffic Theory, Paris, France. p. 139???154.

Fu, H., Liu, N., Hu, G., 2017. Hierarchical perimeter control with guaranteed stability for dynamically coupled heterogeneous urban traffic. Transportation Research Part C: Emerging Technologies 83, 18–38.

Geroliminis, N., Haddad, J., Ramezani, M., 2012. Optimal perimeter control for two urban regions with macroscopic fundamental diagrams: A model predictive approach. IEEE Transactions on Intelligent Transportation Systems 14, 348–359.

de Gier, J., Schadschneider, A., Schmidt, J., Schütz, G.M., 2019. Kardar-parisi-zhang universality of the nagel-schreckenberg model. Physical Review E 100, 052111.

Guo, L., Ye, J.J., 2016. Necessary optimality conditions for optimal control problems with equilibrium constraints. SIAM Journal on Control and Optimization 54, 2710–2733.

Guo, Q., Ban, X.J., 2020. Macroscopic fundamental diagram based perimeter control considering dynamic user equilibrium. Transportation Research Part B: Methodological 136, 87–109.

Haddad, J., 2017. Optimal perimeter control synthesis for two urban regions with aggregate boundary queue dynamics. Transportation Research Part B: Methodological 96, 1–25.

Haddad, J., Ramezani, M., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. Transportation Research Part B: Methodological 54, 17–36.

Hajiahmadi, M., Knoop, V.L., De Schutter, B., Hellendoorn, H., 2013. Optimal dynamic route guidance: A model predictive approach using the macroscopic fundamental diagram, in: 16th international IEEE conference on intelligent transportation systems (ITSC 2013), IEEE. pp. 1022–1028.

Halperin, B., Hohenberg, P., 1969. Scaling laws for dynamic critical phenomena. Physical Review 177, 952.

Hopf, E., 1970. On the right weak solution of the cauchy problem for a quasilinear equation of first order. Indiana Univ. Math. J. 19, 483–487.

Janson, S., 2007. Brownian excursion area, wright's constants in graph enumeration, and other brownian areas. Probability Surveys 4, 80–145.

Keyvan-Ekbatani, M., Yildirimoglu, M., Geroliminis, N., Papageorgiou, M., 2015. Multiple concentric gating traffic control in large-scale urban networks. IEEE Transactions on Intelligent Transportation Systems 16, 2141–2154.

Kobayashi, Y., Takayasu, H., Havlin, S., Takayasu, M., 2021. Robust characterization of multidimensional scaling relations between size measures for business firms. Entropy 23, 168.

Kostinski, S., Amir, A., 2016. An elementary derivation of first and last return times of 1d random walks. American Journal of Physics 84, 57–60.

Kouvelas, A., Saeedmanesh, M., Geroliminis, N., 2017. Enhancing model-based feedback perimeter control with data-driven online adaptive optimization. Transportation Research Part B: Methodological 96, 26–45.

Krause, S.M., Habel, L., Guhr, T., Schreckenberg, M., 2017. The importance of antipersistence for traffic jams. EPL (Europhysics Letters) 118, 38005.

Laval, J., 2021. Power law behavior of elementary cellular automata. Preprints 2021060649. doi:10.20944/preprints202106.0649.v1.

Laval, J.A., Chilukuri, B.R., 2016. Symmetries in the kinematic wave model and a parameter-free representation of traffic flow. Transportation Research Part B: Methodological 89, 168 – 177.

Laval, J.A., Leclercq, L., 2013. The Hamilton-Jacobi partial differential equation and the three representations of traffic flow. Transportation Research Part B 52, 17–30. doi:10.1016/j.trb.2013.02.008.

Lighthill, M.J., Whitham, G.B., 1955. On kinematic waves ii. a theory of traffic flow on long crowded roads. Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 229, 317–345.

Majumdar, S.N., Comtet, A., 2005. Airy distribution function: from the area under a brownian excursion to the maximal height of fluctuating interfaces. Journal of Statistical Physics 119, 777–826.

Mori, T., Smith, T.E., Hsu, W.T., 2020. Common power laws for cities and spatial fractal structures. Proceedings of the National Academy of Sciences 117, 6469–6475.

Nagel, K., 2021. Personal communication.

Nagel, K., Paczuski, M., 1995. Emergent traffic jams. Physical Review E 51, 2909.

Nagel, K., Rasmussen, S., Barrett, C.L., 1996. Network traffic as a self-organized critical phenomena. Technical Report. Los Alamos National Lab., NM (United States).

Nagel, K., Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. Journal de physique I 2, 2221–2229.

Newell, G.F., 2002. A simplified car-following theory : a lower order model. Transportation Research Part B 36, 195–205.

Olmos, L.E., Çolak, S., Shafiei, S., Saberi, M., González, M.C., 2018. Macroscopic dynamics and the collapse of urban traffic. Proceedings of the National Academy of Sciences 115, 12654–12661.

Paczuski, M., Nagel, K., 1996. Self-organized criticality and 1/f noise in traffic, in: Traffic and granular flow, World Scientific Singapore. p. 73.

Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. Transportation Research Part B: Methodological 74, 1–19.

Richards, P.I., 1956. Shock waves on the highway. Operations research 4, 42–51.

Rieser, M., Nagel, K., 2008. Network breakdown "at the edge of chaos" in multi-agent traffic simulations. The European Physical Journal B 63, 321–327.

Schadschneider, A., 2021. Personal communication.

Schroeder, M., 2009. Fractals, chaos, power laws: Minutes from an infinite paradise. Courier Corporation.

Sirmatel, I.I., Geroliminis, N., 2017. Economic model predictive control of large-scale urban road networks via perimeter control and regional route guidance. IEEE Transactions on Intelligent Transportation Systems 19, 1112–1121.

Stauffer, D., Aharony, A., 2018. Introduction to percolation theory. CRC press.

Taleb, N.N., 2020. Statistical consequences of fat tails: Real world preasymptotics, epistemology, and applications. STEM Academic Press.

Wolfram, S., 1984. Cellular automata as models of complexity. Nature 311, 419.

Yang, K., Zheng, N., Menendez, M., 2017. Multi-scale perimeter control approach in a connected-vehicle environment. Transportation research procedia 23, 101–120.

Yildirimoglu, M., Sirmatel, I.I., Geroliminis, N., 2018. Hierarchical control of heterogeneous large-scale urban road networks via path assignment and regional route guidance. Transportation Research Part B: Methodological 118, 106–123.

Zaliapin, I., Kagan, Y.Y., Schoenberg, F.P., 2005. Approximating the distribution of pareto sums. Pure and Applied geophysics 162, 1187–1228.

Zeng, G., Gao, J., Shekhtman, L., Guo, S., Lv, W., Wu, J., Liu, H., Levy, O., Li, D., Gao, Z., et al., 2020. Multiple metastable network states in urban traffic. Proceedings of the National Academy of Sciences 117, 17528–17534.

Zeng, G., Li, D., Guo, S., Gao, L., Gao, Z., Stanley, H.E., Havlin, S., 2019. Switch between critical percolation modes in city traffic dynamics. Proceedings of the National Academy of Sciences 116, 23–28.

Zhang, L., Zeng, G., Li, D., Huang, H.J., Stanley, H.E., Havlin, S., 2019. Scale-free resilience of real traffic jams. Proceedings of the National Academy of Sciences 116, 8673–8678.

Zhong, R., Chen, C., Huang, Y., Sumalee, A., Lam, W., Xu, D., 2018. Robust perimeter control for two urban regions with macroscopic fundamental diagrams: A control-lyapunov function approach. Transportation Research Part B: Methodological 117, 687–707.

Zipf, G., 1949. Human behaviour and the principle of least effort.