


Spatial and temporal memory effects in the Nagel-Schreckenberg model for crowdsourced traffic property determination

Meshkat Botshekan and Franz-Josef Ulm ^{*}*Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

(Received 23 June 2021; accepted 16 September 2021; published 4 October 2021)

We investigate the spatial and temporal memory effects of traffic density and velocity in the Nagel-Schreckenberg cellular automaton model. We show that the two-point correlation function of vehicle occupancy provides access to spatial memory effects, such as headway, and the velocity autocovariance function to temporal memory effects such as traffic relaxation time and traffic compressibility. We develop stochasticity–density plots that permit determination of traffic density and stochasticity from the isotherms of first- and second-order velocity statistics of a randomly selected vehicle. Specifically, provided ergodicity and stationarity, these stochasticity–density plots permit a direct determination of traffic properties from crowdsourced measurements of velocities of vehicles. We illustrate the predictive prowess of the approach for crowdsourced vehicle speed data collected by anonymous smartphone measurements for the state of Massachusetts, USA, as a powerful alternative to classical traffic property estimates from spatially distributed user counts.

DOI: [10.1103/PhysRevE.104.044102](https://doi.org/10.1103/PhysRevE.104.044102)

I. INTRODUCTION

Traffic density, mean velocity, and driver behavior are central mobility indicators amid the backdrop of an ever-increasing demand for reliable navigation systems designed to mitigate intrinsic trade-offs between mobility, safety, and sustainability of our road networks. While most navigation systems approximate traffic density from their user count and average real-time vehicle velocity measurements (obtained spatially by, for instance, smartphones for traffic flux estimates, the so-called floating car data [1]), it has long been recognized that traffic exhibits complex dynamics of a many-body system from free flow to jamming, which defies simple averaging rules [2–6]. Among the many traffic flow models ranging from continuum to agent-driven approaches [7], the Nagel-Schreckenberg (NaSch) cellular automaton model [3] has emerged as a powerful tool not only to reproduce critical features of traffic flow [8], such as backward-moving shock waves [9] and the fundamental diagram of traffic [10], but foremost to track the internal model traffic structure by means of simulations [11–15]. This includes quantitative insights into the mechanism of jamming from investigations of strength and range of interactions between successive vehicles captured by short-range correlation functions [16,17]; phase transition phenomena from investigation of the correlation length in the velocity-velocity covariance function [18]; and jamming rate, jam lifetime, and jam size from stability criteria of the NaSch-model close to the critical jamming density [19]. Key in these studies is the apparent simplicity of the NaSch-model heuristics as a discrete lattice-gas-like model—acceleration

$v_j = \min(v_j + 1, v_{\max})$, deceleration $v_j = \min(d_j, v_j)$ (with d_j the headway), and random deceleration $v_j = \max(v_j - 1, 0)$ with probability p —which permit an update of the j th vehicle position $x_j \rightarrow x_j + v_j$ in units of cells (or integer velocities) in the cellular automaton. In addition to v_{\max} , driver behavior in the NaSch model is condensed into the stochasticity parameter p , bounding the maximum speed of vehicles in free flow. Indicative of the intrinsic unpredictability in driver behavior, the stochasticity parameter drives fluctuations of vehicles' velocity, flux and occupancy, intrinsic to the internal model traffic structure captured by the NaSch-model.

Herein we hypothesize that this internal structure holds critical information relevant for the spatial and temporal mapping of traffic density, mean velocity, and driver behavior from individual driver velocity recordings. We explore this hypothesis by considering the NaSch-model as a stochastic process with intrinsic homogeneity which stems from the analogous collective behavior of drivers and the periodic boundary condition of the system. We demonstrate that, owing to the unique internal structure of the NaSch-model, memory in a vehicle's velocity time history provides the handshake between velocity fluctuations and traffic properties.

The paper is structured as follows: In Sec. II, we ascertain spatial and temporal memory effects in the form of unique scaling relations through the application of, respectively, the two-point correlation function (classically used for texture analysis of porous materials [20,21]) to cell occupancy and the velocity autocovariance function. These scaling relations are at the core of isotherms of first- and second-order velocity statistics for traffic density and driver behavior determination derived in Sec. III. The predictive prowess of the approach is illustrated in Sec. IV through application of the isotherms for traffic property determination from both local detector data and floating car data.

^{*}ulm@mit.edu

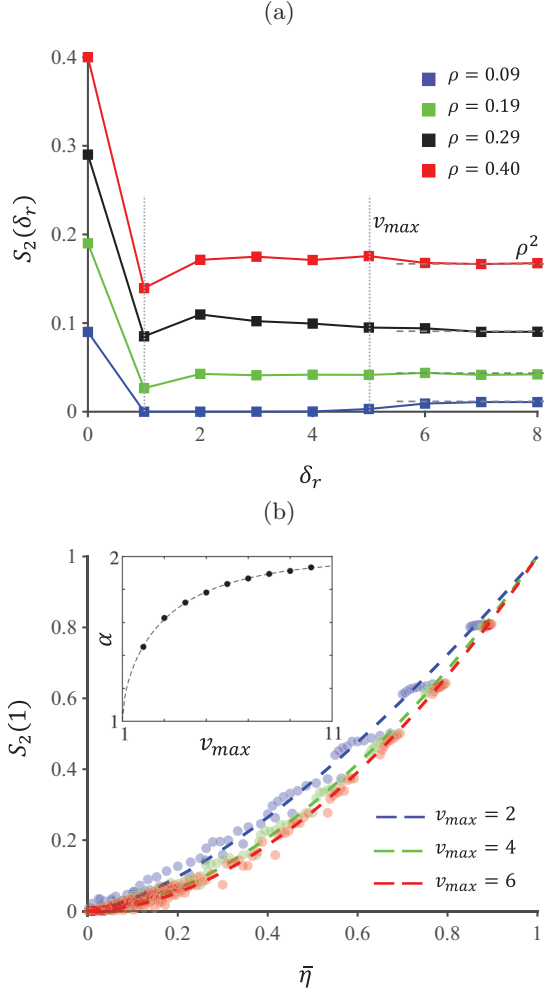


FIG. 1. Two-point correlation function of the NaSch-model representing probability of occupancy: (a) $S_2(\delta_r)$ with asymptotes $S_2(0) = \rho$ and $S_2(\delta_r \rightarrow \infty) = \rho^2$ for $p = 0.5$ and $v_{\max} = 5$; $\rho = 0.09 < \rho_c^p$ is the free-flow regime and shows no correlation with the occupation number of the neighboring cells for $\delta_r \leq v_{\max}$. (b) $S_2(1)$ for the congested flow as a function of $\bar{\eta} = (\rho - \rho_c^p)/(1 - \rho_c^p)$ and $p \in \{0, 0.1, \dots, .9\}$; $S_2(1) \sim$ scales as $S_2(1) \sim \bar{\eta}^\alpha$ with $1 \leq \alpha \approx 1 + \tanh[0.5(v_{\max} - 1)^{0.6}] \leq 2$.

II. SPATIAL AND TEMPORAL MEMORY EFFECTS IN THE NASCH MODEL

A. Two-point correlation function and spatial memory effects

Our starting point is an analysis of two-point correlation [20,21] of cell occupation numbers [Fig. 1(a)],

$$S_2(r_1, r_2) = \mathbb{E}[I_\eta(r_1, t)I_\eta(r_2, t)], \quad (1)$$

where $\mathbb{E}[\cdot]$ and η , respectively, denote the ensemble average operator and jamming random variable, and $I_\eta(r, t) = 1$ if cell r is occupied at time t , otherwise $I_\eta(r, t) = 0$. Given stationarity and ergodicity of the system, the two-point correlation function degenerates to $S_2(r_1, r_2) = S_2(\delta_r = |r_2 - r_1|) = \mathbb{E}[I_\eta(r, t_0)I_\eta(r + \delta_r, t_0)]$, which represents the autocorrelation of a snapshot of the occupation numbers at a given time t_0 . The two-point correlation function $S_2(\delta_r)$ is bound by

[22,23]

$$S_2(\delta_r = 0) = \rho; \quad S_2(\delta_r \rightarrow \infty) = \rho^2, \quad (2)$$

where $\rho = \mathbb{E}[I_\eta(r, t)]$ stands for the traffic density or occupancy. Critical information about the internal traffic structure is provided by the slope of S_2 at the origin in terms of the chord length ℓ_c [22]:

$$\frac{dS_2}{d\delta_r} \Big|_{\delta_r=0} = -\frac{S_2(0)}{\ell_c}. \quad (3)$$

Applied to traffic, the chord length defines the probability of the immediate neighboring cell ($\delta_r = 1$) of an occupied cell to be occupied or not,

$$\mathbb{P}[I_\eta(r+1, t_0) | I_\eta(r, t_0) = 1] = \frac{S_2(1)}{\rho} = 1 - \frac{1}{\ell_c}, \quad (4)$$

where $\mathbb{P}[\cdot]$ stands for the probability operator. In free flow, $\ell_c = 1$, and hence $\mathbb{P}[I_\eta(r+1, t_0) | I_\eta(r, t_0) = 1] = 0$, whereas in complete jamming $S_2(1) = \rho$, and thus $1/\ell_c \rightarrow 0$. In between these asymptotes, a jamming transition occurs at a critical occupancy ρ_c^p . More specifically, when plotted against the probability of the subset $\eta \in N$ to be part of a cluster of jammed vehicles, that is, $\bar{\eta} = \mathbb{P}[\eta \cap \rho > \rho_c^p] \approx (\rho - \rho_c^p)/(1 - \rho_c^p) \in [0, 1]$, we find from simulations that the occupation probability of the next cell neighbor scales as [Fig. 1(b)]

$$S_2(1) = \bar{\eta}^\alpha; \quad \alpha \in [1, 2] \quad (5)$$

The exponent α depends on v_{\max} [inset of Fig. 1(b)] for all v_{\max} and p , when ρ_c^p is estimated from $\rho_c^p = v_J(v_J + \bar{v}_F)^{-1}$ where $\bar{v}_F = v_{\max} - p$ and $v_J = 1 - p$ stand for, respectively, the average free-flow velocity and the analytic upper bound of the jammed dissolution velocity of the NaSch model [24]. Moreover, in the congested regime, we find that $S_2(\delta_r)$ converges rapidly for $\delta_r > 1$ to the asymptotic value $S_2(\delta_r \rightarrow \infty) = \rho^2$ [Fig. 1(a)]. This is indicative of the statistical independence of occupancy of nonadjacent cells, much akin to a Markov process, that is,

$$\mathbb{E}[I_\eta(r, t_0)I_\eta(r + \delta_r, t_0)] \approx \mathbb{E}[I_\eta(r, t_0)]\mathbb{E}[I_\eta(r + \delta_r, t_0)] = \rho^2 \quad (6)$$

for $\delta_r > 1$ and $\bar{\eta} > 0$. This shows that the underlying spatial ergodicity and stationarity of the NaSch-model give rise to traceable *spatial* memory effects, which will turn out to be critical for linking internal structure to traffic flow properties. For instance, taking into account the Markovian property of occupancy I_η , the average headway of a vehicle can be estimated from the two-point correlation function (for derivation, see the Appendix):

$$\mathbb{E}[d] = \bar{d} = \frac{1}{S_2(0)} - 1, \quad (7)$$

with $S_2(0) = \rho$.

B. Velocity autocovariance function and temporal memory effects

A second quantity of interest to ascertain *temporal* memory effects in the internal structure is the velocity autocovariance function, which for stationary processes reduces to

considering the time lag,

$$C_{vv}(\delta\tau = |t_2 - t_1|) = \mathbb{E}[v(t_1, \zeta)v(t_2, \zeta)] - \bar{v}^2, \quad (8)$$

where $v(t, \zeta)$ represents velocity (in NaSch-units) as a stochastic process, while ζ denotes a random event with $\mathbb{P}[\zeta]$ probability of occurrence and $\bar{v} = \mathbb{E}[v(t, \zeta)]$ its mean value. In free flow, the velocity variation of a single vehicle resembles a Bernoulli process of velocity trials v_{\max} and $v_{\max} - 1$ with probabilities $1 - p$ and p , respectively. Thus, the free-flow velocity variance reads $\sigma_{vv,F}^2 = C_{vv,F}(\delta\tau = 0) = p(1 - p)$ and $C_{vv,F}(\delta\tau \neq 0) = 0$ (due to the independence of trials in a Bernoulli process). In contrast, upon jamming [Fig. 2(a)], a distinct time memory effect builds up similar to Ornstein-Uhlenbeck stochastic processes [25–27], which fades exponentially from $C_{vv}(\delta\tau = 0) = \sigma_{vv}^2$ to $C_{vv}(\delta\tau \rightarrow \infty) = 0$. Analogous to the chord length in the two-point correlation function [see Eq. (3)], we capture the time memory effect in the form of a characteristic decay time, τ_c , from the slope of the autocovariance function,

$$\left. \frac{dC_{vv}}{d\delta\tau} \right|_{\delta\tau \rightarrow 0} = -\frac{C_{vv}(0)}{\tau_c}, \quad (9)$$

where $C_{vv}(0) = \sigma_{vv}^2$. The velocity is expected to have the strongest memory upon transitioning from free flow to jammed flow, and to reduce to zero (the so-called memoryless process) for fully jammed traffic. In between these asymptotes, we find from simulations that the decay time, τ_c , scales with the jamming probability, $\bar{\eta} = P(\eta \cap \rho > \rho_c^p)$ [Fig. 2(b)] as $\tau_c \sim \bar{\eta}^{-\gamma_v}$ and can be approximated by

$$\tau_c \approx \hat{\tau}_c(\bar{\eta}^{-\gamma_v} - 1) \quad (10)$$

where $\hat{\tau}_c = 1.88$ and $\gamma_v = 0.56$ are fitting parameters. The τ_c isotherm, which captures all (p, ρ) pairs with same τ_c , reads $p = (1 - \rho(1 + v_{\max}) + v_{\max}\theta_v)/(1 - 2\rho + \theta_v)$ with $0 \leq \theta_v = (\tau_c/\hat{\tau}_c + 1)^{-1/\gamma_v} \leq 1$ and $\theta_v \leq \rho \leq (1 + v_{\max}\theta_v)/(1 + v_{\max})$.

In addition, akin to the kinetic theory of gases, we can define a kinetic compressibility of vehicle speed from the expected value ratio of jammed to free-flow velocity fluctuations [Fig. 2(c)]:

$$\kappa_v = \frac{\mathbb{E}[v^2]}{\mathbb{E}[v_F^2]}; \quad \mathbb{E}[v_F^2] = p(1 - p) + \bar{v}_F^2. \quad (11)$$

The kinetic compressibility decays from $\kappa_v = 1$ in free flow ($\rho \leq \rho_c^p$) to a state of incompressibility, $\kappa_v = 0$ at $\rho = 1$, as a function of the jamming probability, $\kappa_v \approx (1 - \bar{\eta})^\beta$, where the power exponent depends on stochasticity, $\beta = \beta(p)$ [inset of Fig. 2(c)].

C. Local density autocovariance function and temporal memory effects

We proceed by exploiting memory effects in local density $\tilde{\rho}$ defined as the traffic density on a randomly selected subset of adjacent cells of the system $l_{\tilde{\rho}}^{\zeta}$ with a prescribed size of size $|l_{\tilde{\rho}}|$; that is, $\tilde{\rho}(t, \zeta) = \sum_{j \in l_{\tilde{\rho}}^{\zeta}} I_{\eta}(j, t)/|l_{\tilde{\rho}}|$, where $|l_{\tilde{\rho}}|$ is significantly smaller than the system size to be representative of local properties. While the (global) density ρ is a conserved quantity over time, local density is a stochastic

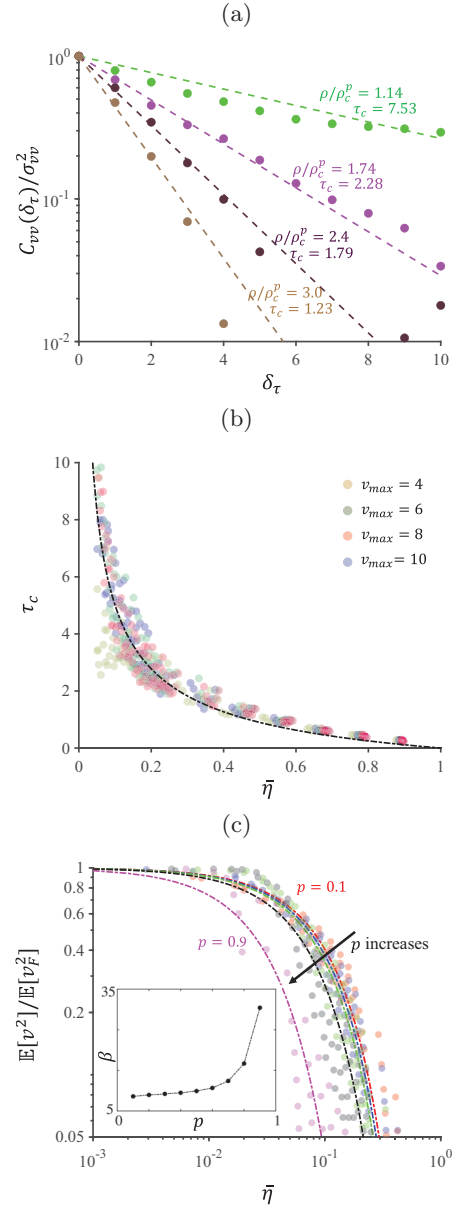


FIG. 2. (a) Velocity autocovariance in NaSch-model decays exponentially, $C_{vv}(\delta\tau) \approx \sigma_{vv}^2 \exp(-\delta\tau/\tau_c)$, resembling an Ornstein-Uhlenbeck process ($v_{\max} = 5, p = 0$). (b) Decay timescale τ_c as a function of $\bar{\eta}$ in the interaction-dominated regime for $p \in \{0.1, 0.2, \dots, 0.9\}$; Velocity approaches memorylessness as traffic density increases with $\tau_c \approx 1.88(\bar{\eta}^{-0.56} - 1)$. (c) Jammed to free-flow velocity fluctuations κ_v for $v_{\max} \in \{4, 6, 8, 10\}$ decays monotonically from 1 (kinetically compressible vehicle speed) at $\bar{\eta} = 0$ to 0 (kinetically incompressible vehicle speed) upon jamming in a power form $\kappa_v \approx (1 - \bar{\eta})^\beta$

process with an expected value of ρ . Given the stationarity of the process, the autocovariance function of local density, $C_{\tilde{\rho}\tilde{\rho}}(|t_1 - t_2| = \delta\tau) = \mathbb{E}[\tilde{\rho}(t_1, \zeta)\tilde{\rho}(t_2, \zeta)] - \rho^2$, adopts a characteristic linear behavior illustrated in Fig. 3(a). The two-point correlation function implies that for high v_{\max} and in the congested regime, $S_2(r, r + \delta_r) \approx \rho^2$ [Sec. II A] indicative of statistical independence between $I_{\eta}(r, t)$ and $I_{\eta}(r + \delta_r, t)$ for $\delta_r > 0$. Therefore, using the Bernoulli assumption of $I_{\eta}(r, t)$,

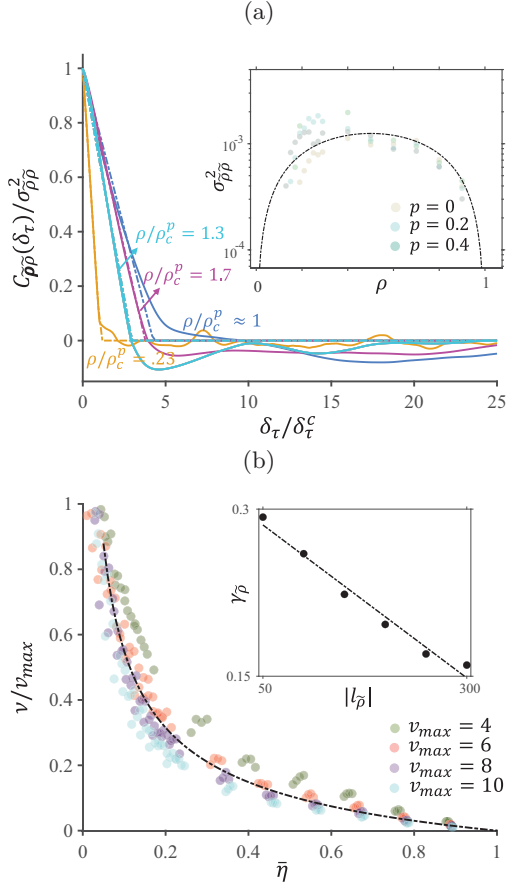


FIG. 3. (a) Normalized autocovariance function of local density exhibiting a distinct linear decay considering $v_{\max} = 5$, $p = 0.3$ and $|l_{\bar{\rho}}| = 100$ for free flow $\rho/\rho_c^p \leq 1$, transitioning to congested flow $\rho/\rho_c^p \approx 1$ and congested flow $\rho/\rho_c^p > 1$; the inset shows the variance of local density $\sigma_{\bar{\rho}}^2$ and the black curve is the variance for Bernoulli process $\sigma_{\bar{\rho}}^2 \approx \rho(1-\rho)/|l_{\bar{\rho}}|$. (b) Universal pattern of v/v_{\max} as a function of jamming probability approximated by $v/v_{\max} \approx 0.5(\bar{\eta}^{-\gamma_v} - 1)$ where $\gamma_v = 5 \times 10^{-4}|l_{\bar{\rho}}| + 0.31$, as shown in the inset.

the variance of local density can be approximated as $\sigma_{\bar{\rho}}^2 \approx \rho(1-\rho)/|l_{\bar{\rho}}|$ [inset of Fig. 3(a)]. The autocovariance function thus permits the following simplification:

$$C_{\bar{\rho}\bar{\rho}}(\delta\tau) \approx \sigma_{\bar{\rho}}^2 \left(1 - \frac{1}{v} \frac{\delta\tau}{\delta\tau^c}\right) \geq 0, \quad (12)$$

where $\delta\tau^c = |l_{\bar{\rho}}|/\bar{v}$ denotes the average residence time of a vehicle traveling on $l_{\bar{\rho}}$ and $v\delta\tau^c$ stipulates the temporal memory of the local density process. In free flow, local density at two time steps are correlated only if their time lag is less than $\delta\tau^c$; and thus $v = 1$. Akin to temporal memory of velocity in the congested flow, memory can be parametrized as a power function of $\bar{\eta}$ [Fig. 3(b)],

$$v \approx v_{\max} \hat{v}(\bar{\eta}^{-\gamma_v} - 1), \quad (13)$$

where $\hat{v} = 0.5$ and $\gamma_v = 5 \times 10^{-4}|l_{\bar{\rho}}| + 0.31$ [inset of Fig. 3(b)] are fitting parameters. Local density shows strong memory upon transitioning to congested flow and approaches memorylessness as $\bar{\eta} \rightarrow 1$.

III. LINK BETWEEN INTERNAL STRUCTURE AND TRAFFIC DENSITY AND STOCHASTICITY

We proceed by matching the found internal traffic structure of spatial and temporal memory effects with traffic density ρ and stochasticity parameter p .

A. Fundamental diagram

To this end, we first construct the fundamental diagram using the results of the two-point correlation function, employing the traffic flux definition $\mathbb{E}[I_{\eta}]\mathbb{E}[v] = \rho\bar{v}$. In the free-flow regime, the independence of cell occupation and vehicle speed provides a linear relation between flux and the average free flow velocity $\bar{v} = \bar{v}_F = v_{\max} - p$:

$$J = \rho\bar{v} = v_F\rho, \quad \rho < \rho_c^p. \quad (14)$$

An equally linear relationship between flux and density approximates flux in jammed flow when considering (i) a first-order estimate of the average velocity—in NaSch-units of number of cells or integer velocity—from the average headway, \bar{d} , in the form $\bar{v} = (1-p)\bar{d}$; and (ii) an estimate of the average headway from the two-point correlation function $\bar{d} = 1/S_2(0) - 1$ [see Eq. (7) and the Appendix], while (iii) assuming the statistical independence of I_{η} and v . That is, letting $S_2(0) = \rho$, the traffic flux–density relation for congested flow is obtained:

$$J = \rho\bar{v} \approx \rho(1-p)\bar{d} = (1-p)(1-\rho), \quad \rho > \rho_c^p. \quad (15)$$

Finally, a combination of Eqs. (14) and (15) leads to the well-known bilinear approximation of the fundamental diagram [28] (cited by Ref. [7]),

$$J = \rho\bar{v} = v_F(\rho - \bar{\eta}), \quad (16)$$

where $\bar{\eta} = |\rho - \rho_c^p|/(1 - \rho_c^p)$. It should be noted that measured flux–density relations typically exhibit large scatter that can be attributed to nonequilibrium traffic conditions. The regularity of the bilinear form we here derive with the help of the two-point correlation function of the NaSch-model relates to its underlying stationarity and ergodicity. In return, given stationarity and ergodicity, the fundamental diagram provides a further relation between a measurable mean velocity \bar{v} , density ρ , and stochasticity p .

B. Stochasticity-density plot

We are now ready to match spatial or temporary memory effects for the determination of traffic density, ρ , and stochasticity parameter, p . From the two-point correlation function, we retain the bilinear flux–density relation (the fundamental diagram), to construct mean velocity isotherms along $p = 1 + \rho\bar{v}/(\rho - 1)$ with $0 \leq \rho \leq 1/(1 + \bar{v})$. In the velocity stochasticity-density (p, ρ) plot [Fig. 4(a)], we overlay these mean velocity isotherms with isotherms of the decay time τ_c as a function of (p, ρ) . Akin to a phase diagram, the curve $0 \leq p = (1 - \rho(1 + v_{\max}))/ (1 - 2\rho) \leq 1$ separates free flow from congested flow. In the congested flow, the \bar{v} and τ_c isotherms intersect at a unique point in the (p, ρ) -plane provided that $\Omega(\bar{v}, \theta_v) = \bar{v}/v_{\max} + (1 + \bar{v})\theta_v \leq 1$ [where $\theta_v = (\tau_c/\hat{\tau}_c + 1)^{-1/\gamma_v}$, with fitting parameters $\hat{\tau}_c = 1.88$ and $\gamma_v = 0.56$, see Fig. 2(b)]. That is, $\Omega(\bar{v}, \theta_v) \leq 1$

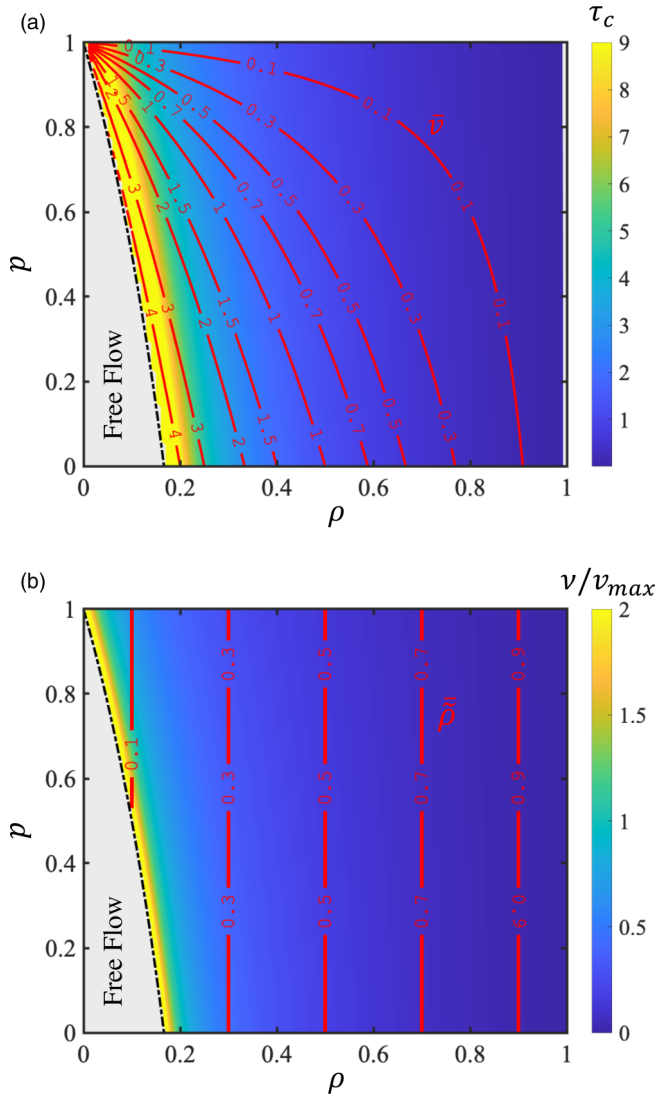


FIG. 4. (a) Velocity and (b) local density stochasticity-density plot showing the interplay between the first-, second-order moments, occupancy ρ , and stochasticity parameter p for $v_{\max} = 5$. Stochasticity-occupancy ($p - \rho$) interaction provides the means to estimate macroscopic traffic properties, p and ρ , from two statistical observables, which are \bar{v} and τ_c [Eq. (10)] for velocity, and $\bar{\rho}$ and v [Eq. (13)] for local density.

is the velocity sample space of the NaSch model indicating the region of all possible \bar{v} and θ_v outcomes. In a similar fashion, we construct the stochasticity-density plot [Fig. 4(b)] for the local density which allows for relating the initial slope of the normalized autocovariance function, $(\sigma_{\bar{\rho}\bar{\rho}}^2)^{-1} dC_{\bar{\rho}\bar{\rho}}/d\delta\tau$, and the expected value of local density, $\mathbb{E}[\bar{\rho}] = \bar{\rho}$, to the stochasticity parameter and traffic density. The v isotherm reads $p = (1 + \theta_v v_{\max} - \rho(1 + v_{\max})) / (1 + \theta_v - 2\rho)$ with $\theta_v = (v / (v_{\max} + 1))^{-1/\gamma_v}$ which intersects the vertical $\bar{\rho}$ isotherms at a unique point if $(\rho(v_{\max} + 1) - 1) / v_{\max} \leq \theta_v \leq \rho$.

In summary, given stationarity and ergodicity, first- and second-order ensemble statistics of velocity or local density, namely, the mean and initial slope of the normalized autocov-

ariance function, provide a means to determine stochasticity parameter and traffic density.

IV. APPLICATION

We now investigate the predictive prowess of our approach for estimating traffic properties from local density measurements and crowdsourced velocity data of individual vehicles. Since our approach relies on ergodicity, we first investigate the conditions under which a realization of the process can statistically represent the ensemble moments. We then employ this condition with empirical local density measurements and crowdsourced velocity data of vehicles.

A. Ergodicity and entropy

Invoking the ergodic theorem of statistical mechanics [29], a vehicle eventually explores the entire phase space in a uniform sense over long (enough) timescales, resulting in an overall ergodic behavior of vehicles in the NaSch model. That is, one realization of the process, say ζ_0 , is statistically rich enough to approximate the ensemble averages from its temporal moments. We apply the ergodic principle to velocity $v(t, \zeta_0)$ of a randomly selected vehicle ζ_0 . That is, recalling the equality of time and phase averages in ergodic mechanical systems, the mean and autocovariance are estimated from

$$\bar{v} \approx \frac{1}{T_r(\zeta_0)} \sum_{t=1}^{T_r(\zeta_0)} v(t, \zeta_0) \quad (17)$$

and

$$C_{vv}(\delta\tau) \approx \frac{1}{T_r(\zeta_0)} \sum_{t=1}^{T_r(\zeta_0)} v(t, \zeta_0)v(t + \delta\tau, \zeta_0) - \bar{v}^2, \quad (18)$$

where $T_r(\zeta_0)$ denotes the representative timescale of realization ζ_0 . This timescale is the shortest time interval over which random event ζ is statistically representative of the ensemble. Focusing on the probability distribution of velocity, such a timescale is controlled by the entropy (also known as the expected information content) of velocity [30]:

$$H(p, \rho) = - \sum_{v_i=0}^{v_{\max}} \mathbb{P}[v_i] \ln(\mathbb{P}[v_i]). \quad (19)$$

The behavior of representative timescale $T_r(\zeta_0)$ with respect to entropy (Fig. 5) is similar to the relaxation time τ_c : it diverges at the transition density [14,15] and increases exponentially as

$$\mathbb{E}[T_r] = T_r^0 \exp(\Gamma \tilde{H}(p, \rho)), \quad (20)$$

where $\tilde{H}(p, \rho) = H(p, \rho) / \ln(v_{\max} + 1)$ is the normalized entropy with $\ln(v_{\max} + 1)$ corresponding to the entropy of a uniform distribution of velocity. While fitting parameter T_r^0 increases linearly as a function of v_{\max} [with $T_r^0 \approx 10(v_{\max} + 1)$ for $2 \leq v_{\max} \leq 10$], the prefactor $\Gamma \approx 2.65$ is independent of v_{\max} . An ensemble with entropy $H(p, \rho)$ is, therefore, expected to be statistically identical to one of its realizations of minimum length $T_r^0 \exp\{\Gamma \tilde{H}(p, \rho)\}$. Equation (20) provides an expected lower bound for achieving ergodicity of velocity in the NaSch model. Similar to velocity, ergodicity in the NaSch model allows us to approximate the ensemble statistics

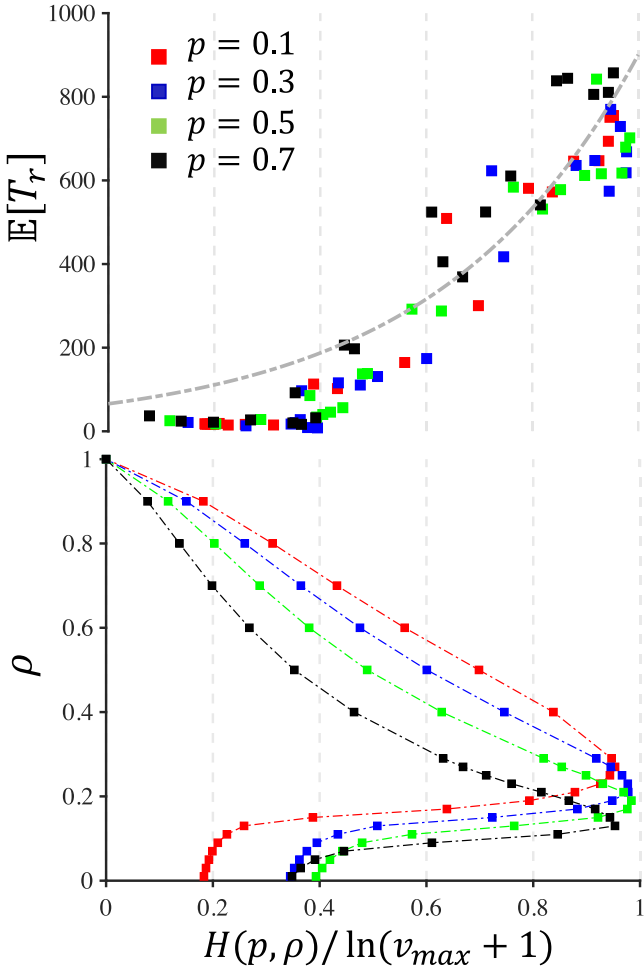


FIG. 5. Variation pattern of entropy $H(p, \rho)$ and its impact on expected representative timescale $\mathbb{E}[T_r]$ for $v_{\max} = 5$. Representative timescale is controlled by entropy and increases as the system becomes more uncertain, corresponding to higher levels of entropy with more possible configurations. Entropy degenerates to $H(p, \rho < \rho_c^p) = -\ln p^p(1-p)^{1-p}$ for the free-flow regime and reaches its upper bound at $\rho \sim \rho_c^p$, where velocity has an almost uniform distribution with $H(p, \rho \sim \rho_c^p) \approx \ln(v_{\max} + 1)$.

of local density from a subset of space when observed over long time intervals.

B. Application to local density data

The first application considers classical traffic density measurements achieved by a fixed sensor (e.g., cameras) along a road. Such measurements provide a means to estimate local density $\bar{\rho}$. The data reported in Ref. [31] were obtained from measurements carried out on the German freeway A1 near an intersection with German freeway A59 in June 1996 [Fig. 6(a)]. Translated in NaSch units, we consider the road capacity N_c (in number of vehicles per hour), which we attribute to the deterministic limit density of the NaSch model, $\rho_c^{p=0} = (v_{\max} + 1)^{-1}$. Denoting by L_c the cell length, a first-order conversion between NaSch units and real (time-length)

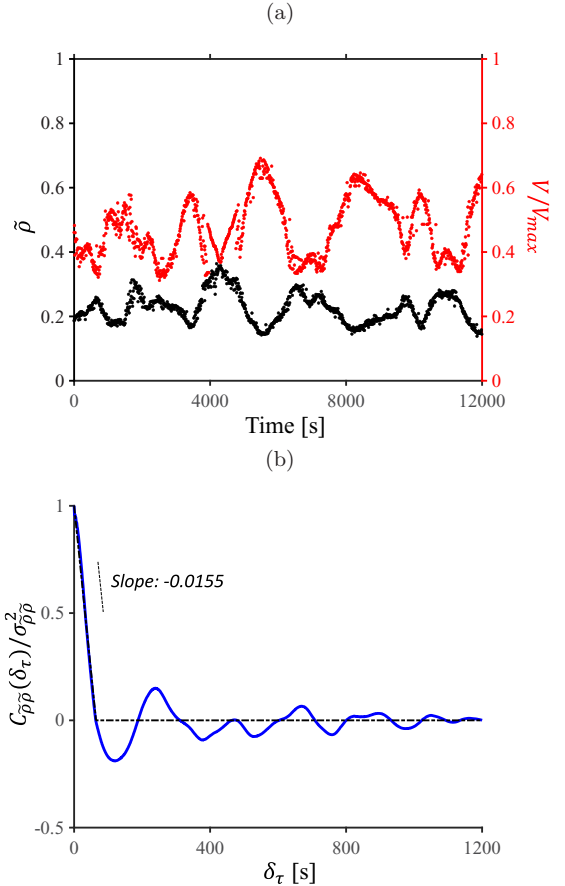


FIG. 6. (a) Summary of measurements adopted from Ref. [31]; $\mathbb{E}[V]/V_{\max} = 0.47$ with $V_{\max} = 120$ km/h and $\mathbb{E}[\bar{\rho}] = 0.22$. (b) The normalized autocovariance function of the local density displays a characteristic linear decay. The statistical properties of the NaSch model allows us to estimate the expected value of velocity $\mathbb{E}[V]$ from the first- and second-order moments of the local density.

units is provided by

$$\frac{V_{\max}/L_c}{(1 + v_{\max})N_c} = \text{const.} \quad (21)$$

For a single-lane road capacity $N_c = 1900$ and for $v_{\max} = 5$, each second corresponds to approximately 1 NaSch time unit considering $L_c \approx 7.5$ meters (as suggested in [3]). The local density shows an expected value of $\rho = \bar{\rho} = 0.22$ and its normalized autocovariance function exhibits a characteristic linear behavior [Fig. 6(b)] with a fitted slope of $-(v\delta_\tau^c)^{-1} = -0.0155$. From the variance of local density $\sigma_{\bar{\rho}}^2 = 2.5 \times 10^{-4}$, we readily recognize from Eq. (12) that $|l_{\bar{\rho}}| \approx 60$, which implies that the local density values are averaged over an approximately 450-meter spatial window. From Eqs. (13) and (15), we obtain the stochasticity parameter $p = 0.3$. Next, from the fundamental diagram [Eq. (15)], we estimate the average velocity, $\bar{v}/v_{\max} = 0.49$. This value which we obtain from the statistical moments of local density is in remarkable agreement with the average velocity obtained from the recorded measurements, i.e., $\mathbb{E}[V]/V_{\max} = 0.47 \approx \bar{v}/v_{\max} = 0.49$. This shows that the internal structure provides an independent means to estimate not only average traffic speed,

but an estimate of the stochasticity parameter reminiscent of driver behavior as well.

C. Application to crowdsourced vehicle velocity data

The second application we consider illustrates the use of our approach for crowdsourced velocity data. In contrast to the fixed sensor application, sensors are installed in moving vehicles. The vehicle speed data were collected by anonymous users on main roads in the Commonwealth of Massachusetts, USA, through the Carbin educational app [32,33], which records the velocity from the GPS position at a 1 Hz frequency. To avoid interference with road regulations such as traffic lights, school zones, speed bumps, etc., we focus on data recorded on roads of speed limits $V_{\max} \geq 45$ mph (≥ 72.4 km/h). As drivers are observed to drive 10% to 20% faster than the posted speed limit in free flow, this speed limit can be considered a real-life lower speed limit of free-flow traffic speed. The conversion of velocity measurements into NaSch-units is thus performed via $v = \lfloor v_{\max}(V/V_{\max}) + 0.5 \rfloor \leq v_{\max}$, where $\lfloor \cdot \rfloor$ denotes the floor operator. We consider this conversion in our analysis of more than 31 000 miles of speed measurements acquired over a time span of one year by anonymous users covering almost the entire main road network of Massachusetts. The length of the time window, T_m , was checked *a priori* to satisfy the ergodicity condition [Eq. (20)] which is an underlying assumption for our analysis, that is, $T_m - T_r^0 \exp(\Gamma \bar{H}) \leq 0$ with \bar{H} denoting the normalized entropy of velocity distribution over time window of T_m . Furthermore, the results of the analysis were checked *a posteriori* to satisfy the NaSch event space condition, i.e., $\Omega(\bar{v}, \theta_v) \leq 1$ and, moreover, $\kappa_v \approx (1 - \bar{\eta})^\beta$. By way of example, Fig. 7 displays the analysis of (a) a sample velocity measurement $V(t)/V_{\max}$ of a 20-minute trip on interstate highway I-95 together with (b) its conversion into NaSch (velocity) units v , (c) evolution of traffic density ρ and transition density ρ_c^p predicted from the velocity profile, and (d) examples of autocovariance functions at densities higher and close to the transition density, showing the intimate interplay of decay time with traffic density.

The so-obtained results were partitioned into four time intervals: (i) 6:00 to 10:00, (ii) 10:00 to 15:00, (iii) 15:00 to 19:00, and (iv) 19:00 to 00:00, where in each time interval there are at least 10^3 analysis results. By taking into account the probability of observing memoryless velocity profiles given $v_{\max} - \bar{v} \leq 1$, i.e., $\mathbb{P}[\tau_c \approx 0 \mid v_{\max} - \bar{v} \leq 1]$, we find that the free-flow probability for time intervals (i) and (iii) is around 10%, whereas for time intervals (ii) and (iv) it increases to almost 15%. This is in agreement with the average weekday daily traffic data reported by the Boston Metropolitan Planning Organization [34]. Figure 8 depicts the geospatial distribution of traffic density and stochasticity parameter for time interval (iii). It is found that the expected traffic density and stochasticity parameter around the urban area of Boston (inset in Fig. 8) is, respectively, 1.3 and 1.15 times the one of rural area (the region outside the red box), implying higher average velocity in rural areas. The network-level expected traffic densities are $\mathbb{E}[\rho] = [0.12, 0.11, 0.13, 0.10]$, and expected stochasticity parameters are $\mathbb{E}[p] = [0.7, 0.63, 0.67, 0.62]$ for time intervals

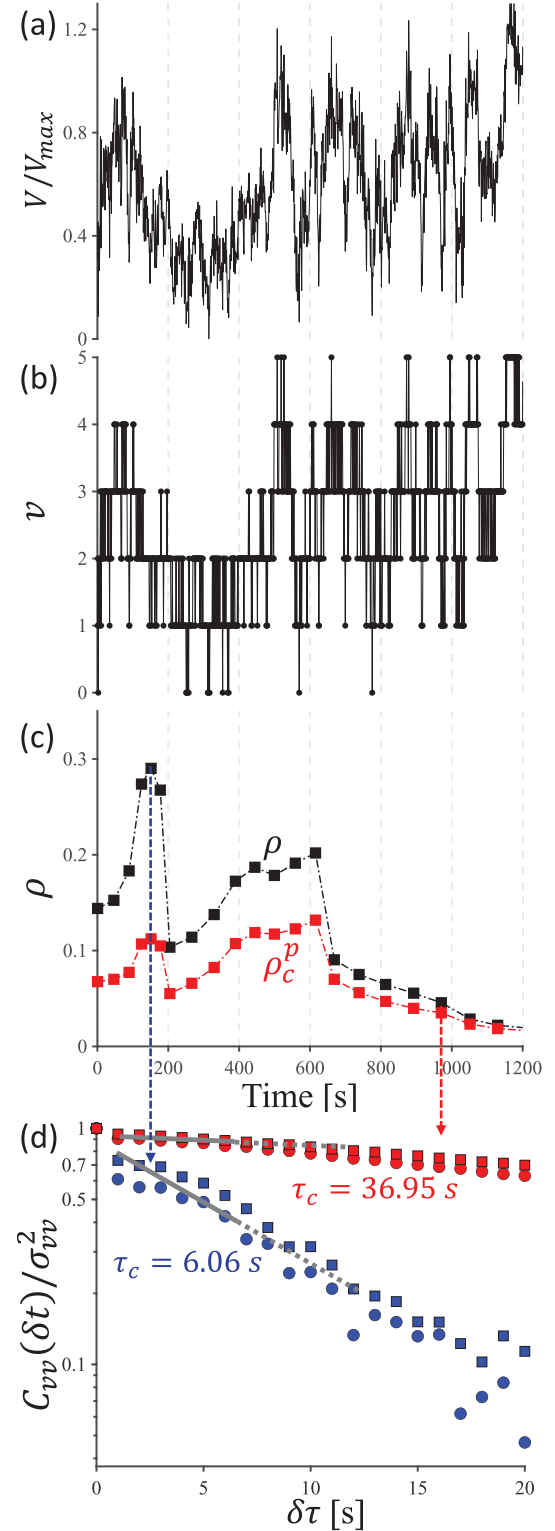


FIG. 7. Sample analysis: (a) Velocity time history of a vehicle driving on I-95, a north-south interstate highway in Massachusetts, USA, (b) NaSch representation of velocity profile, (c) evolution of inferred traffic density ρ and transition density ρ_c^p , and (d) autocovariance functions at two times (squares and circles represent the autocovariance of velocity measurement and its NaSch representation, respectively). The initial slope of the autocovariance function is inversely proportional to τ_c [Eq. (9)]; memory of velocity signal is shorter at the higher density.

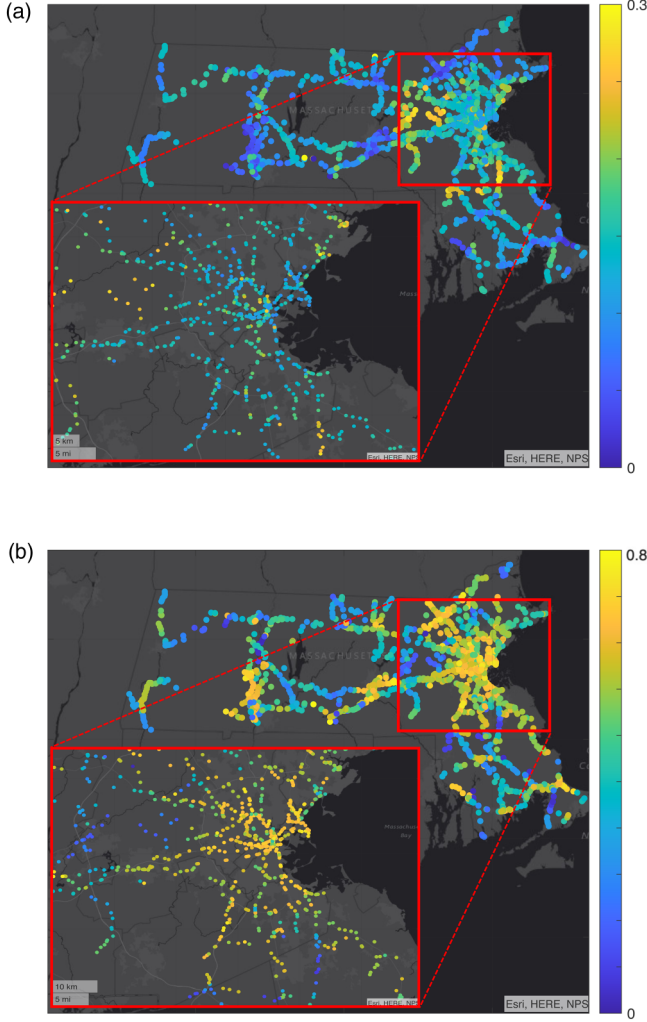


FIG. 8. Geospatial distribution of (a) traffic density ρ and (b) stochasticity parameter p for the time interval 15:00 to 19:00, on main roads in Massachusetts, USA, determined from crowdsourced 1 Hz vehicle velocity recordings (data collected over a 12-month period with Carbin educational app [32,33]).

(i)–(iv). The traffic density and stochasticity parameter in time intervals (i) and (iii) are higher than in time intervals (ii) and (iv). This suggests that drivers show a more erroneous driving behavior during rush hours. Furthermore, the inferred traffic parameters imply that, in an average sense, traffic is predominantly in the congested flow regime.

V. CONCLUSIONS

In summary, we have shown that spatial and temporary memory effects expressed by second-order moments of occupancy and velocity hold critical information relevant for the spatial and temporal mapping of traffic density and driver behavior that can be assessed from individual driver velocity recordings provided ergodicity and stationary. In particular, the two-point correlation function of occupancy provides access to spatial memory effects, such as headway, whereas the velocity autocovariance function provides access to temporal memory effects in the form of the decay time and

traffic compressibility. Taken together, the isotherms provide a means to access traffic density and stochasticity from density-stochasticity plots. The fact that these higher statistical moments are directly accessible by crowdsourced velocity measurements provides a powerful alternative to classical traffic property estimates from spatially distributed user counts. Finally, it should be noted that the theory and model calibration herein presented are restricted to single-lane traffic. Beyond this paper, the approach merits extensions to multiple lane models of the NaSch type with lane-change probabilities [35–37], ramp exits [38,39], and corrections for traffic obstacles for inner city applications.

ACKNOWLEDGMENTS

Research carried out by the Concrete Sustainability Hub (CSHub@MIT), with funding provided by the Portland Cement Association (PCA) and the Ready Mixed Concrete Research & Education Foundation (RMC E&F).

APPENDIX: AVERAGE HEADWAY ESTIMATION FROM TWO-POINT CORRELATION FUNCTION

We aim at deriving the average headway from probability considerations. Our starting point is the probability q_{ij} of a cell $r + 1$ being in a state i (occupied or empty) conditioned by state j of cell r :

$$\begin{aligned} q_{ij} &= \mathbb{P}[I_\eta(r+1) = i \mid I_\eta(r) = j] \\ &= \frac{\mathbb{P}[I_\eta(r+1) = i \cap I_\eta(r) = j]}{\mathbb{P}[I_\eta(r) = j]}, \end{aligned} \quad (\text{A1})$$

where $\mathbb{P}[A \cap B]$ stands for the joint probability. Hence, the conditional probability of a cell $r + 1$ to be occupied ($i = 1$) given that cell r is occupied ($j = 1$) is readily obtained when recognizing from Eq. (1) that $\mathbb{P}[I_\eta(r+1) = i \cap I_\eta(r) = j] = \mathbb{E}[I_\eta(r+1)I_\eta(r)] = S_2(1)$ and $\mathbb{P}[I_\eta(r) = 1] = S_2(0)$, hence

$$q_{11} = \mathbb{P}[I_\eta(r+1) = 1 \mid I_\eta(r) = 1] = \frac{S_2(1)}{S_2(0)}. \quad (\text{A2})$$

Given the binary nature of occupation, Eq. (A2) allows us to determine the probability of cell $r + 1$ being empty when cell r is occupied:

$$q_{01} = 1 - q_{11} = \frac{S_2(0) - S_2(1)}{S_2(0)}. \quad (\text{A3})$$

Since $\mathbb{P}[A|B] = \mathbb{P}[A \cap B]/\mathbb{P}[B]$ and $\mathbb{P}[B|A] = \mathbb{P}[B \cap A]/\mathbb{P}[A]$, we readily derive from the expression of q_{10} the probability that cell $r + 1$ is occupied when cell r is empty; that is,

$$q_{10} = q_{01} \frac{\mathbb{P}[I_\eta(r) = 1]}{\mathbb{P}[I_\eta(r) = 0]} = q_{01} \frac{S_2(0)}{1 - S_2(0)}. \quad (\text{A4})$$

Finally, the conditional probability of two cells, $r + 1$ and r , being empty is obtained from

$$q_{00} = 1 - q_{10} = \frac{1 - 2S_2(0) + S_2(1)}{1 - S_2(0)}. \quad (\text{A5})$$

With the probabilities in hand, we can determine the average headway while making use of the found Markov property:

$$\mathbb{P}[I_\eta(r + \delta_r) | I_\eta(r)] = \mathbb{P}[I_\eta(r + \delta_r)], \quad \delta_r > 1. \quad (\text{A6})$$

For illustration, consider a realization I_η of the form

$$1 - 0 - 0 - 1 \quad (\text{A7})$$

(i.e., cells 1 and 4 are occupied while cells 2 and 3 are empty). From the Markov property (A6), we know that occupancy of cell 3 (respectively, 4) is independent of cell 1 occupancy (respectively, 1 and 2). Otherwise said, headway probability reduces to the pairs 1 – 0 (cells 1 and 2), 0 – 0 (cells 2 and 3), and 0 – 1 (cells 3 and 4) defined by probabilities q_{01} , q_{00} , and q_{10} . The probability of observing the realization

1 – 0 – 0 – 1, which is the headway $d = 2$, is thus

$$q_{01} \times q_{00} \times q_{10} = \frac{(S_2(0) - S_2(1))^2 (1 - 2S_2(0) + S_2(1))}{S_2(0)(1 - S_2(0))^2}. \quad (\text{A8})$$

To generalize, consider that the probability of observing a headway d is $q_{10} \times q_{01} \times q_{00}^{d-1}$; whence the average headway:

$$\bar{d} = q_{10} \times q_{01} \times \sum_{d=1}^{\infty} d \times q_{00}^{d-1} = \frac{1}{S_2(0)} - 1, \quad (\text{A9})$$

where we used the geometric series development,

$$\sum_{d=1}^{\infty} d \times q_{00}^{d-1} = \left(\sum_{d=1}^{\infty} q_{00}^d \right)' \Big|_{0 \leq q_{00} \leq 1} \stackrel{=}{=} \frac{1}{(1 - q_{00})^2}, \quad (\text{A10})$$

with $()'$ denoting derivation with respect to q_{00} .

-
- [1] T. Jeske, Proc. of the BlackHat Europe, 1 (2013).
- [2] M. J. Lighthill and G. B. Whitham, *Proc. R. Soc. London A* **229**, 317 (1955).
- [3] K. Nagel and M. Schreckenberg, *J. Phys. I* **2**, 2221 (1992).
- [4] H. Y. Lee, H.-W. Lee, and D. Kim, *Phys. Rev. Lett.* **81**, 1130 (1998).
- [5] J. Matsukidaira and K. Nishinari, *Phys. Rev. Lett.* **90**, 088701 (2003).
- [6] H. M. Zhang, *Transp. Res. Part B Methodol.* **36**, 275 (2002).
- [7] F. van Wageningen-Kessels, H. van Lint, K. Vuik, and S. Hoogendoorn, *EURO J Transp Logist* **4**, 445 (2015).
- [8] M. Schreckenberg, A. Schadschneider, K. Nagel, and N. Ito, *Phys. Rev. E* **51**, 2939 (1995).
- [9] N. Geroliminis and C. F. Daganzo, *Transp. Res. Part B Methodol.* **42**, 759 (2008).
- [10] F. L. Hall, B. L. Allen, and M. A. Gunter, *Transp. Res. Part A: Gen.* **20**, 197 (1986).
- [11] L. Roters, S. Lübeck, and K. D. Usadel, *Phys. Rev. E* **59**, 2672 (1999).
- [12] N. Boccara and H. Fuks, *J. Phys. A: Math. Gen.* **33**, 3407 (2000).
- [13] S.-p. Chen and D.-w. Huang, *Phys. Rev. E* **63**, 036110 (2001).
- [14] M. Sasvári and J. Kertész, *Phys. Rev. E* **56**, 4104 (1997).
- [15] B. Eisenblätter, L. Santen, A. Schadschneider, and M. Schreckenberg, *Phys. Rev. E* **57**, 1309 (1998).
- [16] S. Cheybani, J. Kertész, and M. Schreckenberg, *J. Phys. A: Math. Gen.* **31**, 9787 (1998).
- [17] N. Lakouari, K. Jetto, H. Ez-Zahraouy, and A. Benyoussef, *Int. J. Mod. Phys. C* **25**, 1350089 (2014).
- [18] N. Bain, T. Emig, F.-J. Ulm, and M. Schreckenberg, *Phys. Rev. E* **93**, 022305 (2016).
- [19] H. M. Bette, L. Habel, T. Emig, and M. Schreckenberg, *Phys. Rev. E* **95**, 012311 (2017).
- [20] S. Torquato, J. Beasley, and Y. Chiew, *J. Chem. Phys.* **88**, 6540 (1988).
- [21] Y. Jiao, F. Stillinger, and S. Torquato, *Proc. Natl. Acad. Sci.* **106**, 17634 (2009).
- [22] S. Torquato, *Appl. Mech. Rev.* **44**, 37 (1991).
- [23] P. Smith and S. Torquato, *J. Comput. Phys.* **76**, 176 (1988).
- [24] M. Gerwinski and J. Krug, *Phys. Rev. E* **60**, 188 (1999).
- [25] D. T. Gillespie, *Phys. Rev. E* **54**, 2084 (1996).
- [26] F. Castro, A. D. Sánchez, and H. S. Wio, *Phys. Rev. Lett.* **75**, 1691 (1995).
- [27] J. Łuczka, P. Hänggi, and A. Gadomski, *Phys. Rev. E* **51**, 2933 (1995).
- [28] C. Daganzo, *Transp. Res. Part B Methodol.* **28**, 269 (1994).
- [29] C. C. Moore, *Proc. Natl. Acad. Sci.* **112**, 1907 (2015).
- [30] T. M. Cover, *Elements of Information Theory* (John Wiley & Sons, New Jersey, 1999).
- [31] K. Nagel, P. Wagner, and R. Woesler, *Oper. Res.* **51**, 681 (2003).
- [32] Carbin Educational Smartphone Application (Android) (2021), <https://play.google.com/store/apps/details?id=com.carbin.carbin2>.
- [33] Carbin Educational Smartphone Application (iOS) (2021), <https://apps.apple.com/us/app/carbin/id1559648286/>.
- [34] Traffic Volumes, Boston Region MPO (2021), <https://www.ctps.org/subjects/traffic-volumes/>.
- [35] M. Rickert, K. Nagel, M. Schreckenberg, and A. Latour, *Physica A* **231**, 534 (1996).
- [36] K. Nagel, D. E. Wolf, P. Wagner, and P. Simon, *Phys. Rev. E* **58**, 1425 (1998).
- [37] X.-G. Li, B. Jia, Z.-Y. Gao, and R. Jiang, *Physica A* **367**, 479 (2006).
- [38] R. Jiang, Q.-S. Wu, and B.-H. Wang, *Phys. Rev. E* **66**, 036104 (2002).
- [39] G. Diedrich, L. Santen, A. Schadschneider, and J. Zittartz, *Int. J. Mod. Phys. C* **11**, 335 (2000).