Counting the corners of a random walk and its application to traffic flow

Florian Knorr and Michael Schreckenberg

Fakultät für Physik, Universität Duisburg-Essen, 47048 Duisburg, Germany

E-mail: `SURNAME@ptt.uni-due.de`

Abstract. We study a system with two types of interacting particles on a one-dimensional lattice. Particles of the first type, which we call "active", are able to detect particles of the second type (called "passive"). By relating the problem to a discrete random walk in one dimension with a fixed number of steps we determine the fraction of active and detected particles for both open and periodic boundary conditions as well as for the case where passive particles interact with both or only one neighbors. In the random walk picture, where the two particles types stand for steps in opposite directions, passive particles are detected whenever the resulting path has a corner. For open boundary conditions, it turns out that a simple mean field approximation reproduces the exact result if the particles interact with one neighbor only.

A practical application of this problem is heterogeneous traffic flow with communicating and non-communicating vehicles. In this context communicating vehicles can be thought of as active particles which can by front (and rear) sensors detect the vehicle ahead (and behind) although these vehicles do not actively share information.

Therefore, we also present simulation results which show the validity of our analysis for real traffic flow.

PACS numbers: 02.10.Ox,05.50.+q, 05.40.Fb, 45.70.Vn

Submitted to: *J. Phys. A: Math. Gen.*

1. Introduction

A combinatorial problem of heterogeneous traffic flow with communicating and non-communicating vehicles originally motivated our analysis: Information exchange between vehicles is expected to enhance traffic safety and traffic stability notably [1, 2, 3, 4, 5, 6]. The envisioned technology for this purpose is wireless communication networks, so-called vehicular ad hoc networks (VANETs). After market introduction, however, only a small fraction of vehicles will be equipped with the necessary communication devices. On the other hand, local sensors (e.g., radar or lidar) enable vehicles to gather information about the preceding and succeeding vehicle. Consequently, a communicating vehicle may not only broadcast information about itself but also information about the vehicle driving behind or ahead. Hence, the number of vehicles whose position is known may be significantly higher than the actual number of communicating vehicles.

Not only for traffic related applications it is an interesting question how many vehicles are on average known (i.e., either actively communicating or being detected by a neighboring vehicle) for a given share of communicating vehicles on the road.

To answer this question we study a one-dimensional particle chain with two types of particles to which we will either refer as "active" or "passive". The traffic flow example in mind also explains our choice to speak of active and passive particles. When studying exclusion processes [7], one would rather speak of occupied and empty sites and in the context of zero range processes (e.g., [8]) of sites and particles, respectively.

For random configurations of active and passive particles we determine the average number of passive particles neighboring an active one by mapping the system to a one-dimensional (1d) discrete random walk. This mapping even allows for an interpretation which is closer to physics: If one thinks of the different particle types as changes in the height profile of a 1d surface by ± 1, then the fraction of passive particles next to an active one corresponds to the surface's extremal-point density. In general, such extremal-point densities allow to study the dynamics of nonequilibrium surface fluctuations and have applications to an even broader range of research. For a detailed discussion of the dynamics of rough surfaces, the density of local extrema and analytic solutions we refer to the very comprehensive article [9] by Toroczkai et al.

The remainder of the paper is organized as follows: After introducing the problem in the next section we will then derive analytic formulas to determine the fraction of known particles depending on the number of active particles for both open and periodic boundary conditions. We will also examine the case where a passive particle becomes known only if its right (left) neighbor is active. In the picture of vehicular traffic this corresponds to the case where communicating vehicles are equipped with front (rear) sensors only. It turns out that in this case the mean field approximation reproduces the exact result. Finally, we compare the analytic results with traffic flow simulations. We find that on a macroscopic level the theoretical findings are in excellent agreement with the simulation results. From a single vehicle's perspective, though, the limited communication range and interference effects decrease the fraction of known vehicles compared to the theoretical prediction.

2. Model description

To analyze the problem presented in the introduction we start with a one-dimensional lattice with N sites. Each lattice site contains either an active or a passive particle (represented by $\bullet$ or $\circ$). Let the number of active (passive) particles be A (P) and $N = P + A$. Then it follows directly the probability of a randomly selected site to contain an active (passive) particle is $p = A/N$ $(1 - p = P/N)$.

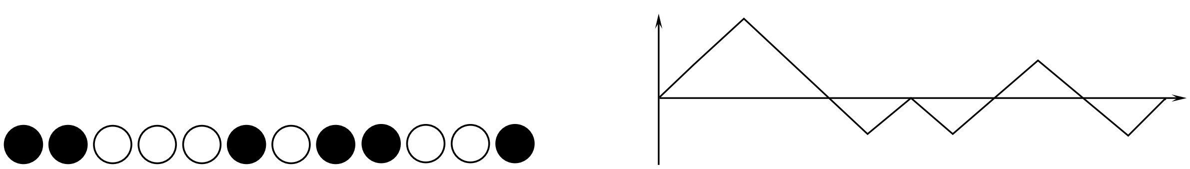
Active particles are always assumed to be known. For a passive particle to be known it has to neighbor an active one. Here we distinguish two cases:

(i) **Symmetric case**: A passive particle is called known if at least one of its two neighbors is active.

(ii) **Asymmetric case**: A passive particle is called known only if its left (right) neighbor is active.

The knowledge about the systems obviously depends on the average number of passive particles which are known in a given configuration. Fig. 1(a) depicts a sample configuration with $N/2 = A = P = 6$. In this configuration all but one (three) particle(s) are known in the symmetric (asymmetric) case independent of the boundary condition.

In general, the number of known particles depends on the distribution and amount of active particles. For instance, to have full knowledge of the system (i.e., for all particles to be known) the minimum share of active particles is $p = 1/3$ in the symmetric case ($\circ\bullet\circ\circ\bullet\circ$) and $p = 1/2$ in the asymmetric case ($\bullet\circ\bullet\circ$).



(a) Particle chain (b) Corresponding random walk

Figure 1. The presented model consists of a chain with two types of particles: active ones and passive ones. 1(a) Active (passive) particles are represented by $\bullet(\circ)$. The number of known particles consists of all active particles and all passive particles next to an active one. In the sample configuration of 1(a) there are 11 particles out of 12 known in the symmetric case and 9 out of 12 in the asymmetric case. The model can also be viewed as a 1D random walk where the different types of particles correspond to steps in opposite directions. In this picture passive particles become known whenever the direction of the walk is reversed. Hence, the fraction of known passive particles is also related to the density of extremal points of the corresponding random walk.

At first sight, the usage of only two particle types, representing communicating and non-communicating vehicles, appears as a rough approximation of real traffic flow as we ignore the inter-vehicle distance. We admit the introduction of additional empty lattice sites would better reflect the actual situation on a real road or in traffic cellular automata (e.g., [10]) as considered in section 4. But the approximation is justified since the detection range of vehicle sensors of about 200 meters is larger than the typical space gap between two vehicles. Besides, the case where the inter-vehicle spacing exceeds the sensors' range is negligible because in this case the contribution of vehicular communication to traffic safety and efficiency will be marginal at best.

3. Determining the fraction of known particles

In this section we determine the fraction of known particles $k_{\mathrm{a,s}}^{\mathrm{o,p}}$ as a function of the total particle number N and the number of active particles A. The upper index in $k_{\mathrm{a,s}}^{\mathrm{o,p}}$ specifies whether we consider open (o) or periodic (p) boundaries. The lower index denotes the symmetric (s) or asymmetric (a) case.

As a passive particle becomes known when it is next to an active one, one has to count all $\bullet\circ$ (and $\circ\bullet$) sequences among all $\binom{N}{A}$ possible configurations in the asymmetric (symmetric) case. We will start to derive an analytic solution for the symmetric case for periodic and open boundary conditions. The extension to the asymmetric case is straight forward. For the latter case it actually turns out that a simple mean field approximation is exact. The mean field approximation can be represented by a single-valued function $k_{\mathrm{a,s}}(p)$ with $p = A/N$.

3.1. Analytic solution for the symmetric case

It is helpful to interpret the occupied lattice as a one-dimensional random walk where each active particle corresponds to a step to the left and each passive particle to a step to the right or vice versa. Fig. 1(b) illustrates this analogy. In this picture passive particles are discovered whenever the direction of the walk is reversed (i.e., when a corner occurs). Care must be taken, however, when a passive particle neighbors two active particles ($\bullet\circ\bullet$): In this case only one of the two corresponding corners stands for a newly discovered particle.

Hence, the $k_{\mathrm{s}}^{\mathrm{o}}(N, A)$ can be calculated by determining the number of paths with R corners and weighing it by a factor proportional to R. Thereby, our model is similar to a problem by Feynman and Hibbs [11]: To calculate the kernel of a relativistic particle moving in $1 + 1$ dimension Feynman suggested summing over all possible paths with A (P) steps to the left (right) and weighing each path with a factor depending on the number of corners.

To count the configurations which have exactly R corners we have to distinguish four cases:

(i) The first particle is active and the last one is passive ($\bullet \ldots \circ$).

(ii) The first particle is passive and the last one is active ($\circ \ldots \bullet$).

(iii) Both the first particle and last particle are active ($\bullet \ldots \bullet$).

(iv) Neither the first particle nor the last particle are active ($\circ\ldots\circ$).

In analogy to Jacobson and Schulman who presented a solution [12] to Feynman's problem we introduce the function $\Phi_{xy}(R)$. The function returns the number of configurations with R corners if the first particle is of type x and the last one is of type y. For $x = \bullet$ and $y = \circ$ one obtains [12]:

$$\Phi_{\bullet\circ}(R) = \binom{A-1}{\frac{1}{2}(R-1)}\binom{P-1}{\frac{1}{2}(R-1)} \tag{1}$$

with odd R and $0 \le R-1 \le 2\min(A-1, P-1)$. Due to the problem's symmetry it immediately follows $\Phi_{\bullet\circ}(R) = \Phi_{\circ\bullet}(R)$.

For $\Phi_{\circ\circ}(R)$ we obtain:

$$\Phi_{\circ\circ}(R) = \binom{P-1}{\frac{1}{2}R-1}\binom{A-1}{\frac{1}{2}R} \tag{2}$$

for even R and $2 \le R \le \min(2A-2, 2P)$ from which follows

$$\Phi_{\bullet\bullet}(R) = \binom{A-1}{\frac{1}{2}R-1}\binom{P-1}{\frac{1}{2}R} \tag{3}$$

with R even and $2 \le R \le \min(2P-2, 2A)$.

To count passive particles in a $\bullet\circ\bullet$ configuration only once, one of the corresponding corners must be skipped. With periodic boundaries the probability $p_{\bullet\circ\bullet}(N, A)$ for such a configuration to occur on a lattice with N sites and A active particles is (see, e.g., [13])

$$p_{\bullet\circ\bullet}(N, A) = \frac{A-1}{N-2}. \tag{4}$$

For convenience, we introduce

$$n = N\binom{N}{A}.$$

Then it directly follows that the average fraction of known particles $k_{\mathrm{s}}^{\mathrm{o}}$ as a function of N and A is:

$$k_{\mathrm{s}}^{\mathrm{o}}(N, A) = \frac{1}{n}\sum_{R}\sum_{x,y\in\{\circ,\bullet\}}\Phi_{xy}(R)\left(R + A - \frac{R}{2}p_{\bullet\circ\bullet}(N+1, A)\right) \tag{5}$$

where we assumed the $\Phi_{xy}(R)$ to return zero if R is not in the set of valid values as specified in equations (1)–(3). The $1/n$-term normalizes the probability such that $k(N, N) = 1$. The $+1$-term in the first argument of $p_{\bullet\circ\bullet}(\cdot,\cdot)$ is a consequence of the open boundary conditions: As with open boundaries the leftmost and rightmost particles have only one neighbor site, the probability for a $\bullet\circ\bullet$ sequence to occur is identical to the one on a periodic lattice with $N+1$ sites where the additional site is occupied by a passive particle. By using the normalization $\sum_R\sum_{x,y\in\{\bullet,\circ\}}\Phi_{x,y}(R) = \binom{N}{A}$ we can rewrite equation (5) as

$$k_{\mathrm{s}}^{\mathrm{o}}(N, A) = \frac{A}{N} + \left(\frac{1}{n} - \frac{1}{2n}\frac{A-1}{N-1}\right)\sum_{R}\sum_{x,y\in\{\circ,\bullet\}} R \times \Phi_{xy}(R). \tag{6}$$

The sums in equation (6) can be converted to hypergeometric functions [14] and turn out to be particular cases of Gauss's hypergeometric theorem. Thereby, the previous formula can be reduced to a simple rational function depending only on the total number of particles and the number of active particles:

$$k_{\mathrm{s}}^{\mathrm{o}}(N, A) = \frac{A(A^2 + A - 3NA + 3N^2 - 2N)}{N^2(N-1)}. \tag{7}$$

The extension to periodic boundaries is straight forward. When closing the chain to form a ring, one additional corner originates for all $\Phi_{\bullet\circ}(R)$ and $\Phi_{\bullet\circ}(R)$ configurations. In this case one has to replace $R \to R + 1$ for the weighing factor in round brackets in equation (5) and $\bullet\circ\bullet$ sequences need no special treatment which leads to

$$k_{\mathrm{s}}^{\mathrm{p}}(N, A) = \frac{A(A^2 + 3A - 3AN + 3N^2 + 2 - 6N)}{N(N-1)(N-2)}. \tag{8}$$

Finally, let us compare the error of a simple mean field approximation with the exact result. One obtains the mean field approximation by noting that the only configuration in which a passive particle remains unknown is when both its neighbors are passive particles as well. The probability for this configuration is $(1-p)^3$. In all other cases $(1 - (1-p)^3)$ a passive particle is known which yields

$$k_{\mathrm{s}}(p) = k_{\mathrm{s}}(A/N) = p^3 + 3p(1-p) \tag{9}$$

The larger is the system (i.e., the number of lattice sites N) the better is the agreement between equation (7)/(8) and (9) as in the thermodynamic limit $N \to \infty$ and $p = \mathrm{const.}$ (i.e., $N \approx (N-1)$ and $A^i/N^j \to 0$ for $0 < i < j$ and $A < N$), the mean field approximation converges to $k(N, A)$ as $k(N, A) = k(p) + O(N^{-1})$.

3.2. Analytic solution for the asymmetric case

In the asymmetric case a passive particles becomes known only if its left neighbor is active. (We restrict the discussion to this case, although the results are also valid if a passive particle requires an active particle to its right in order to be known.)

With similar reasoning as in the previous section one can derive the fraction of known particles as a function of lattice size N and the number of active particles A. In general, every second corner in the corresponding random walk stands for a known passive particle. Care must be taken for the $\Phi_{xy}(R)$ configurations with $x \neq y$: These configurations have an odd number of corners. Moreover, there is at least one known passive particle for each $\Phi_{\bullet\circ}(R)$ configuration which is not guaranteed for $\Phi_{\bullet\circ}(R)$ configurations. This leads to

$$\begin{aligned} k_{\mathrm{a}}^{\mathrm{o}}(N, A) = \frac{1}{n} \sum_{R} \Bigg[& (\Phi_{\bullet\bullet}(R) + \Phi_{\circ\circ}(R)) \left(\frac{R}{2} + A\right) \\ & + \Phi_{\bullet\circ}(R) \left(\frac{R+1}{2} + A\right) \\ & + \Phi_{\bullet\circ}(R) \left(\frac{R-1}{2} + A\right) \Bigg]. \end{aligned} \tag{10}$$

Evaluating the sums gives

$$k_{\mathrm{a}}^{\mathrm{o}}(N,A) = k_{\mathrm{a}}^{\mathrm{o}}(p) = p(2-p) \tag{11}$$

for open boundaries. This is exactly the same result a mean field approximation yields as one can easily verify: In the asymmetric case the only configuration in which a passive particle remains unknown is if the particle to its left is passive (○○). The probability for this sequence is $(1-p)^2$ and, consequently, $k(p) = 1-(1-p)^2 = p(2-p)$.

A similar calculation for periodic boundary conditions yields:

$$k_{\mathrm{a}}^{\mathrm{p}}(N,A) = \frac{A(2N-A-1)}{N(N-1)}. \tag{12}$$

3.3. The asymmetric case and the density of local maxima

We would like to discuss briefly the relation between the fraction of known passive particles and the density of local maxima in the random walker's path, which we have already mentioned in the introductory section. This relation becomes obvious by comparing figs. 1(a) and 1(b): Each •○ configuration in fig. 1(a) results in a local maximum in fig. 1(b). Hence, we obtain the average density of local maxima $\rho_{\mathrm{max}}^{\mathrm{o}}$ by setting $A=0$ on the rhs of (10):

$$\rho_{\mathrm{max}}^{\mathrm{o}}(N,A) = \frac{A(N-A)}{N^2} = p(1-p). \tag{13}$$

As we average over all configurations, the density of local maxima and minima is identical ($\rho_{\mathrm{max}}^{\mathrm{o}} = \rho_{\mathrm{min}}^{\mathrm{o}}$). An analogous calculation for periodic boundary conditions yields

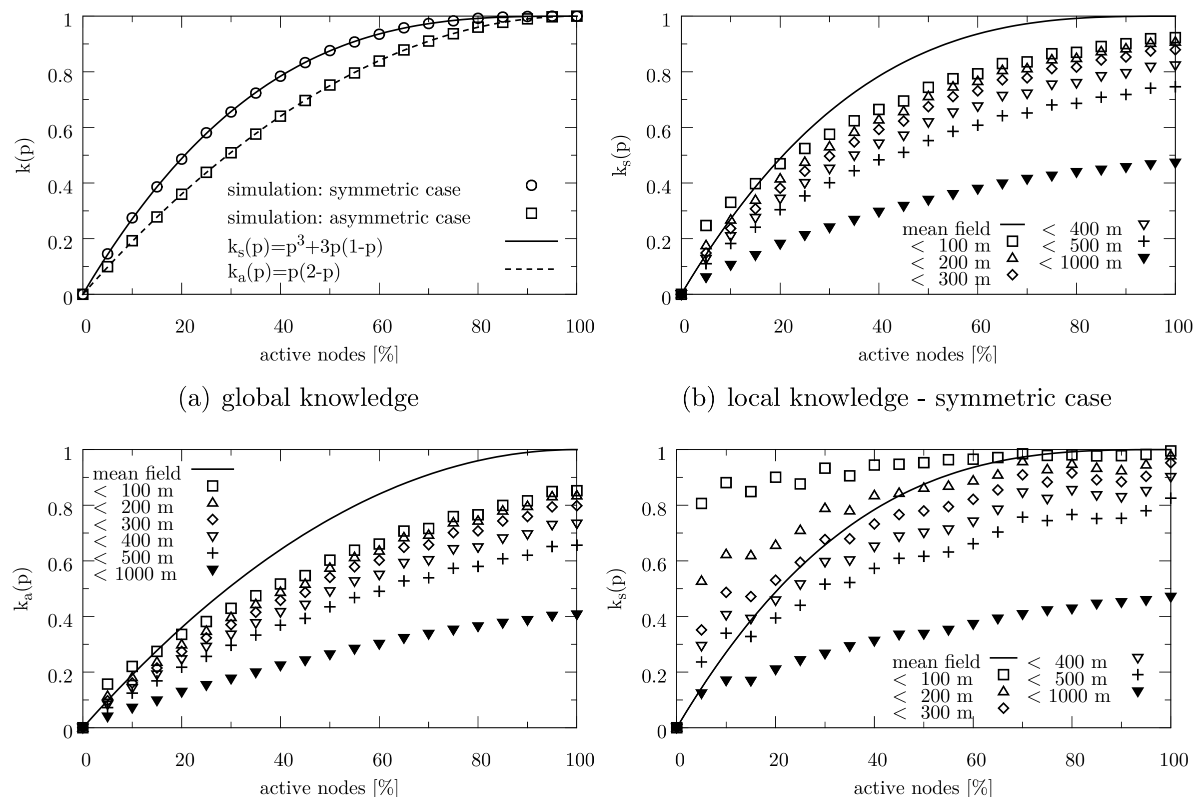
$$\rho_{\mathrm{max}}^{\mathrm{p}}(N,A) = \rho_{\mathrm{min}}^{\mathrm{p}}(N,A) = \frac{A(N-A)}{N(N-1)}. \tag{14}$$

(Note that the interpretation of periodic boundary conditions for the random walker's path is somewhat difficult as the two ends of the 1d surface have different heights if $A \neq P$.)

4. Simulations for vehicular networks

Finally, we validate the theoretical predictions with simulations of a vehicular communication network on a circular one-lane road of radius 1.5 km. Each communicating vehicle is assumed to send periodic status messages with a frequency of 4 Hz (The full set of parameters can be found in [15]). These messages comprise the vehicle's position and velocity as well as the position and the velocity of the preceding (and following) vehicle(s) in the asymmetric (symmetric) case. For realistic communication modeling we used a probabilistic propagation model [16]. Vehicle motion is simulated using a traffic cellular automaton [17] . The road was initialized with densities ranging from 5 percent up to 65 percent. Higher densities were omitted because at a density of 65 percent the average bumper to bumper distance already is below 4.1 meters and a large traffic jam spans the entire road. Thus, from a practical point of view it is not necessary to know the position and velocity of each vehicle to estimate the traffic dynamics at such high vehicle densities. For our analysis we initialized the road homogeneously and recorded the communication statistics of a 60 second interval after a relaxation time of 240 seconds. For each density we averaged over at least five independent runs.

To assess the validity of the theoretical calculations for real world applications we compare the simulations results to the corresponding mean field approximation. Here we distinguish between global and local knowledge. Global knowledge is the aggregated information from all communicating vehicles at a given time. A central node or sever to which all vehicles send the available information might possess such global knowledge. Analogously, we refer to the knowledge of a single vehicle as local knowledge.



(a) global knowledge

(b) local knowledge - symmetric case

(c) local knowledge - asymmetric case

(d) local knowledge at fixed vehicle density $\rho = 9\%$ - symmetric case

Figure 2. We compare the theoretical prediction of $k(p)$ to simulations of a vehicular communication network. Assuming all communicating vehicles successfully transmit their knowledge to a central node the information of the latter depending on the fraction of communicating vehicles is very well described by functions $k(N, A)$ derived in section 3 (fig. 2(a)): The curve for the (a)symmetric case is depicted by a (dashed) solid line. Simulation results were averaged over various vehicle densities in the range $5\% \leq \rho \leq 65\%$ for rates of communicating vehicles sampled over 5% intervals. The average fraction of known vehicles for a single vehicle is given in fig. 2(b)-2(d) for different sizes of the neighborhood.

In fig. 2(a) the percentage of known vehicles of is shown for the entire system. The functional relations (9) and (11) serve as reference. The good agreement is not surprising as we assumed each vehicle is able to successfully transmit its knowledge to the central node.

A realistic treatment of communication between particles, however, is likely to deteriorate these results for two reasons. The communication range is limited and in the case of simultaneous transmissions closer senders are preferred due to a stronger signal strength at the receiver. Hence, the fraction of known particles will decrease the larger is the distance two vehicles. Figs. 2(b) and 2(c) show a comparison between the average knowledge of single vehicles and the predictions by equations (9) and (11), and illustrate this effect. We divided neighboring vehicles in different categories containing all vehicles closer than $100, 200, ..., 500$ and 1000 meters, respectively. For each range we determine the average knowledge a communicating vehicle has in the symmetric (fig. 2(b)) as well as in the asymmetric case (fig. 2(c)). Even if all vehicles are communicating, the fraction of known vehicles stays below 93 percent. For low rates of communicating vehicles the number of known vehicles is above the analytic prediction, though. To explain this behavior see fig. 2(d) where we averaged only over a fixed vehicle density of $\rho = 9\%$ (symmetric case). At low densities there are only few vehicles within the neighborhood of communicating vehicles among which at least the ones immediately in front and behind are detected. This explains why these configurations show a relatively large fraction of known vehicles and, thereby, increase the fraction of known vehicles when averaging over all densities as in fig. 2(b).

5. Discussion

In this paper we developed a model of a one-dimensional system with two types of particles, namely active and passive particles. By mapping the model to a one-dimensional random walk we could derive formulas to determine fraction of known particles within the system depending on the system size N and the number of active particles A. As expected, the functions are strictly increasing for $A \to N$. For low fractions $p = A/N$ of active particles the dominating term in equations (7) and (8) is $3A/N$ for the symmetric case. This in turn means each active particles discovers approximately two passive particles. For $p = 0.1$ and $N \to \infty$ the number of discovered passive particles per active is slightly above 1.7 as can be obtained from equation (9). Similarly, with an asymmetry in particle discovery the term $2A/N$ contributes most to the sums in equations (11) and (12) for $A \ll N$. In the latter case we could also show that a simple mean field approximation is exact for open boundaries. In all cases mean field approximations are in good agreement with the exact result for large N ($N > 100$) as the error decreases linearly with N^{-1}. Furthermore, in systems with open boundaries knowledge is lower than in systems with periodic boundaries as the active particles occupying the boundaries have only one neighbor to discover instead of two.

Comparing the analytic results to simulations of a vehicular communication network which motivated our analysis showed that realistic communication modeling decreases the fraction of known vehicles. The analytic solution then can serve as an upper boundary in most situations.

Acknowledgments

We thank A Schadschneider for discussing selected aspects of the problem with us. We also thank the anonymous reviewer who pointed us to the relation between the problem considered here and the density of local extrema of a 1d surface (cf. [9]). The work of FK was funded by the state of North Rhine-Westphalia and the European Union within the NRW-EU Ziel 2 program.

References

[1] Car 2 car communication consortium. http://www.car-to-car.org/.

[2] Kerner B S, Klenov S L, and Brakemeier A. In *IEEE Intell. Vehicles Symp.*, pages 180 – 185, 2008.

[3] Kerner B S. *Introduction to Modern Traffic Flow Theory and Control: The Long Road to Three-Phase Traffic Theory*, pages 162–168. Springer, Berlin, 2009.

[4] Kerner B S, Klenov S L, and Brakemeier A. *Traf. Eng. & Cont.*, 51(6):217–222, 2010.

[5] Lee H K and Kim B J. *Physica A*, 390(23–24):4555 – 4561, 2011.

[6] Knorr F and Schreckenberg M. *Physica A*, 391(6):2225 – 2231, 2012.

[7] Derrida B and Evans M R. *Nonequilibrium Statistical Mechanics in One Dimension*, chapter 14, pages 277–304. Camebridge University Press, Camebridge, 1997.

[8] Evans M R. Phase transitions in one-dimensional nonequilibrium systems. *Braz. J. Phys.*, 30(1):42–57, 2000.

[9] Toroczkai Z, Korniss G, Das Sarma S, and Zia R K P. *Phys. Rev. E*, 62:276–294, 2000.

[10] Nagel K and Schreckenberg M. *J. Phys. France I*, 2(12):2221–2229, 1992.

[11] Feynman R P and Hibbs A R. *Quantum Mechanics and Path Integrals*, pages 34–36. McGraw-Hill, New York, 1965.

[12] Jacobson T and Schulman L S. *J. Phys. A*, 17:375–383, 1984.

[13] Pulkkinen O and Merikoski J. *Phys. Rev. E*, 64:056114, Oct 2001.

[14] Koepf W. *Hypergeometric summation.* Vieweg, Wiesbaden, 1998.

[15] Knorr F, Baselt D, Schreckenberg M, and Mauve M. *submitted to IEEE Trans. Veh. Technol.*, 2012.

[16] Nakagami M. The m-distribution - a general formual of intensitivity distribution of rapid fading. In W. G. Hoffman, editor, *Statistical Methods in Radio Wave Propagation*, pages 3–36. Pergamon, Oxford, UK, 1960.

[17] Knospe W, Santen L, Schadschneider A, and Schreckenberg M. *J. Phys. A*, 35(15):3369–3388, 2002.