Multiple Particle Correlation Analysis of Many-Particle Systems: Formalism and Application to Active Matter

Rüdiger Kürsten,[1*] Sven Stroteich,[1] Martín Zumaya Hernández,[2,3] and Thomas Ihle[1]
[1]*Institut für Physik, Universität Greifswald, Felix-Hausdorff-Str. 6, 17489 Greifswald, Germany*
[2]*Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México, Apartado Postal 48-3, Código Postal 62251, Cuernavaca, Morelos, Mexico*
[3]*Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Código Postal 04510, Ciudad de México, Mexico*

(Received 9 October 2019; accepted 5 February 2020; published 26 February 2020)

We introduce a fast spatial point pattern analysis technique that is suitable for systems of many identical particles giving rise to multiparticle correlations up to arbitrary order. The obtained correlation parameters allow us to quantify the quality of mean field assumptions or theories that incorporate correlations of limited order. We study the Vicsek model of self-propelled particles and create a correlation map marking the required correlation order for each point in phase space incorporating up to ten-particle correlations. We find that multiparticle correlations are important even in a large part of the disordered phase. Furthermore, the two-particle correlation parameter serves as an excellent order parameter to locate both phase transitions of the system, whereas two different order parameters were required before.

DOI: 10.1103/PhysRevLett.124.088002

In statistical physics, many-particle systems can be studied either by numerical simulations or with analytical theories. Most analytical calculations are based on approximations such as the mean field approach which is among the most widely applied techniques both in and out of equilibrium. To improve mean field theories, low order correlations between two or more particles can be taken into account. In equilibrium, and if only two-particle interactions are involved, the knowledge of the pair correlation function is already sufficient to obtain the equation of state. However, higher order correlations affect quantities such as the heat capacity [1]. Pair correlations have been considered, e.g., in the ring-kinetic theories of Refs. [2–4]. Other authors considered also third and higher order correlations, for example, Refs. [5–9]. An alternative approach are Minkowski functionals [10] that contain information on correlations of all orders.

In real systems, correlations of all orders build up due to interactions. One aim of this Letter is to provide an efficient technique for experimental and simulation data to quantify the validity of assumptions on multiparticle correlations, not only two- and three-particle correlations but up to arbitrary order.

In many cases, there is an intrinsic length scale R associated with the interactions between particles. It is intuitive to study correlations on that length scale because only the nearby particles directly influence the dynamics of a given particle. Hence, it is a minimum requirement for an accurate theory to reproduce the correct number of neighbors that reside within distance R around a given particle. Clearly, this neighbor number is a random variable. We call its probability measure neighbor distribution (ND).

In this Letter, we extract the minimal amount of information from the multiparticle correlation functions that is necessary to reproduce the correct ND. We find that each correlation order contributes exactly two parameters to the parametrization of the ND. Efficient techniques are developed to sample these parameters. We derive the *exact* ND by taking into account multiparticle correlations of finite but arbitrary order. Comparing the calculated distribution to the measured one, we obtain a lower bound on the correlation order that is required for an accurate description of the system.

We apply the analysis to a two-dimensional model of aligning self-propelled particles, the Vicsek model [11,12], which is known to exhibit two different phase transitions [13–16]. We measure a quantitative correlation map of the model by marking the minimal required correlation order for each point in phase space. It is observed that even in the disordered phase multiparticle correlations are important in a large parameter range and the mean field assumption is valid only for very large noise strengths. In addition, we find that a two-particle correlation parameter serves as an excellent order parameter to investigate both phase transitions, whereas in the past two different order parameters have been used [15].

The technique is applicable to arbitrary many-particle systems. However, it is in particular useful out of equilibrium, since there, equilibrium thermodynamics is not available. We expect to detect all phase transitions that go along with spatial rearrangements (e.g., liquid-gas transitions) with the help of a single local correlation parameter. Furthermore, the technique might be employed in the analysis of spatial point patterns [17–19] with applications in diverse fields such as ecology [20,21], neuroanatomy [22], astronomy [23], cell biology [24], or image analysis [25].

In the following we assume that the system is spatially homogeneous and consists of many ($N \to \infty$) identical particles. First, we ask for the number of particles within distance R around an arbitrary fixed point in space which is in general not coinciding with any particle position. If the particles are statistically independent, this number is Poisson distributed [26] with mean $C_1 = B_d(R)\rho$, where $\rho = N/V$ is the density and $B_d(R)$ the volume of a ball of radius R in d dimensions.

Relaxing the assumption of statistical independence and taking two-particle but no higher order correlations into account, we find the probability of s particles to be within distance R to an arbitrary fixed point as [27]

$$p(s) = (C_1 - C_2)^s \exp(C_2/2 - C_1) \times \sum_{k=0}^{\infty} \left[\frac{C_2}{2(C_1 - C_2)^2}\right]^k \frac{1}{k!(s-2k)!}, \quad (1)$$

where we use the convention that $1/l! = 0$ for $l < 0$. Except for the mean number of particles in a ball, C_1, this distribution depends also on the correlation parameter C_2 where C_k in general is defined by

$$C_k := N^k \int G_k(1, \ldots, k) \prod_{l=1}^{k} \theta(R - |\mathbf{r}_l|) dl, \quad (2)$$

where θ represents the Heaviside function. Equation (2) reminds us of the integrals over Mayer functions that appear in the virial expansion of hard spheres [28–30]. Here however, we integrate over a single graph (star) instead of all biconnected graphs and use the k-particle correlation function G_k (also called Ursell functions [31]) as weight. They are defined by a cluster expansion [32]. For example, the first two correlation functions are $G_1(1) := P_1(1)$, $G_2(1,2) := P_2(1,2) - P_1(1)P_1(2)$. Here, the argument "1" represents all degrees of freedom of particle 1 and so on, and P_1 and P_2 are the one- and two-particle probability density functions, respectively. Note that G_2 is related to the radial distribution function as $G_2(\mathbf{r}_1, \mathbf{r}_2) = [g(|\mathbf{r}_2 - \mathbf{r}_1|) - 1]/V^2$. We explicitly calculate the characteristic function of the probability distribution (1) and generalize it by incorporating not only two-particle correlations but correlations up to an arbitrary but finite order $l_{\max}$ to obtain [27]

$$\chi(u) = \exp\left[\sum_{l=1}^{l_{\max}} \sum_{t=0}^{l} (-1)^{l+t} \frac{C_l}{l!} \binom{l}{t} \exp(itu)\right]. \quad (3)$$

Note that $C_2 = C_3 = \ldots = C_{l_{\max}} = 0$ yields the characteristic function of a Poisson distribution with mean C_1, which we recover also by setting $C_2 = 0$ directly in Eq. (1). Hence we call $\chi(u)$ from Eq. (3) the characteristic function of the correlation-induced generalized Poisson distribution. To the best of our knowledge, the distribution defined by Eq. (3) has not been reported in the literature.

In the mean field scenario the ND equals the distribution of the number of particles within an arbitrarily located ball. However, in the correlated case these distributions are different, and the characteristic function of the ND, the probability of an arbitrary particle to have k neighbors $p_n(k)$, is related to $\chi(u)$ given in Eq. (3) via

$$\chi_n(u) = \sum_{l=1}^{l_{\max}} \sum_{t=0}^{l-1} \binom{l-1}{t} (-1)^{t+l+1} \frac{1}{(l-1)!} \times D_l \exp(itu)\chi(u). \quad (4)$$

We give a detailed derivation of Eqs. (3) and (4) in the Supplemental Material [27] and Ref. [33]. Here, the correlation coefficients D_k are defined by

$$D_k := N^{k-1} \int G_k(1,2,\ldots,k) d1 \prod_{l=2}^{k} \theta(R - |\mathbf{r}_1 - \mathbf{r}_l|) dl, \quad (5)$$

where $D_1 := 1$. Thus the correlation coefficients C_k and D_k are the desired minimal set of numbers necessary to calculate the correct ND.

We demonstrate how to extract these parameters from experimental or simulation data. In fact, they depend on two types of directly measurable quantities [27]. The first quantity $V_o(l)$ is the overlap volume of balls of radius R drawn around each particle of an l-plet, summed over all ordered l-plets and normalized by the total system volume, see Supplemental Material [27] for details. The second quantity μ_l is obtained as the expectation value of $s!/(s-l)!$, where s is the number of neighbors of a randomly selected particle. Hence, μ_l can be sampled by counting the number of neighbors for each particle. The overlap can be either sampled by directly calculating the overlap volume of all l-plets or via an effective Monte Carlo algorithm [27] that is independent on the spatial dimension. For the direct calculation in two dimensions we derive formulas for the overlap area of two and three [34] circles depending only on the distances between the centers of the circles [27]. From those sampled quantities the relevant correlation parameters are obtained as [27]

$$C_l = V_o(l) - \sum_{k=1}^{l-1} C_k V_o(l-k) \binom{l-1}{k-1}, \quad (6)$$

$$D_l = \mu_{l-1} - \sum_{k=1}^{l-1} D_k V_o(l-k) \binom{l-1}{k-1}. \quad (7)$$

The above theory is a powerful tool to systematically analyze multi particle correlations of arbitrary order on the relevant length scale in many-particle systems. Once the

correlation coefficients C_k and D_k have been sampled, the ND can be calculated taking into account correlations up to the desired order. Then, one can compare this calculated distribution to the directly measured one. Very good agreement would indicate that higher correlations might be neglected. However, if the calculated and measured ND totally disagree it is certain that higher order correlations are profoundly affecting the dynamics of the system and hence also its steady state.

We apply the correlation analysis to one of the prototypes of active matter, the standard Vicsek model in two dimensions; see Supplemental Material [27] for the definition of the model. Additionally, we apply the analysis technique to a continuous time variant of the Vicsek model that is closely related to direct experimental applications [36]. The results of this second model are qualitatively equivalent to the ones of the Vicsek model and are shown in the Supplemental Material [27].

In Fig. 1 we present an example of the ND from a molecular dynamics simulation. The red circles represent the directly measured distribution. The other symbols show distributions that have been calculated according to Eq. (4) taking into account various correlation orders. We see that the Poisson distribution as well as the distribution that incorporates two-particle correlations do not agree very well with the measured distribution. By additionally taking into account three-particle correlations the distribution is already quite close to the measured ND, and with up to seven-particle correlations it agrees almost perfectly. We use the Kullback-Leibler divergence (KL) [37] to quantify the agreement between measured and calculated distribution. The logarithm of the KL is shown in the lower panel of Fig. 1 when different correlations orders are considered. The agreement improves gradually until seven-particle correlations are considered. For higher order correlations we obtain no improvement, which can be understood looking at the accuracy with which the correlation parameters have been measured. In Table I we display the values of the correlation parameters with standard deviation of the mean in brackets for the same simulation as shown in Fig. 1. Up to seventh order the standard deviation is less then 1% whereas it significantly increases for the next orders. It is possible to measure even higher order correlation parameters accurately by analyzing significantly more data.

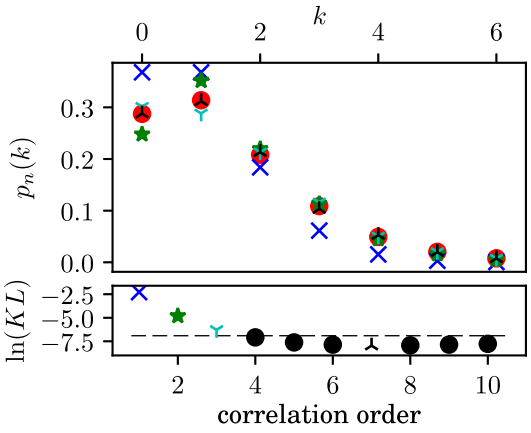

FIG. 1. Top: Neighbor distribution for a randomly picked particle in the standard Vicsek model. Red circles display the measured distribution, blue crosses the Poisson distribution, green stars, cyan Y-shaped markers and black mirrored Y-shaped markers represent the distribution given by Eqs. (3), (4) including up to two-particle, up to three-particle, and up to seven-particle correlations, respectively. System parameters: average neighbor number $C_1 = 1$, noise strength $\eta = 0.48$, free path $v_0\tau = 1$, interaction radius $R = 1$, particle number $N = 22\,500$. Averages have been calculated over 24 realizations and 10^6 time steps after a thermalization period of 10^5 time steps for each realization. Bottom: Logarithm of the Kullback-Leibler divergence between measured ND and calculated one using correlations up to different orders.

To quantify the minimum order of correlations required to accurately reproduce the correct ND we set a threshold for the KL of 10^{-3} which is also displayed in the bottom part of Fig. 1 as a dashed line. We consider the agreement as good if the KL is less than the threshold and as unsatisfying otherwise. Using this definition we determine the minimal order of correlations that need to be considered. The results of extensive simulations for different values of noise strength and particle density are shown in Fig. 2 presenting a quantitative correlation map of the standard Vicsek model. We observe that a mean field hypothesis, as it is usually assumed when deriving field theories like, e.g., Refs. [38,39], is valid only for very large noise strengths. For smaller noise we need two-, three-, four-particle, or even higher correlations. Certainly, the required correlation order depends on the arbitrarily chosen threshold for the KL. However, reasonable modifications of the threshold do not change the overall picture, see Supplemental Material [27].

TABLE I. Correlation parameters measured in the same simulation as the one displayed in Fig. 1 with standard deviation in brackets.

Correlation order k	C_k	D_k
1	1	1
2	0.350 211 5(92)	0.434 489 7(60)
3	0.230 994(21)	0.297 475(21)
4	0.236 633(53)	0.307 351(80)
5	0.344 41(19)	0.44878(31)
6	0.6655(12)	0.8678(16)
7	1.616(11)	2.102(10)
8	4.70(12)	6.121(91)
9	15.7(1.1)	20.69(86)
10	59(10)	78.0(8.1)

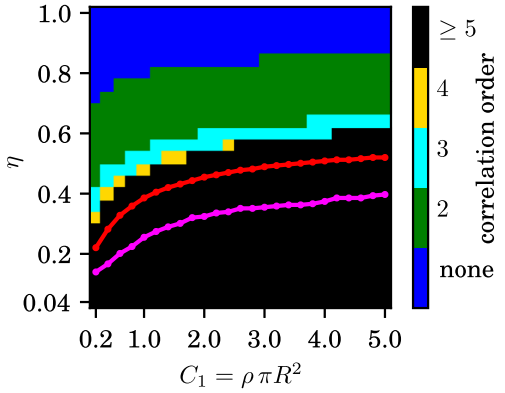

FIG. 2. Correlation map providing the minimal order of correlations necessary to reproduce the measured neighbor distribution up to a Kullback-Leibler divergence of 10^{-3} for the standard Vicsek model depending on noise strength η and particle density. The transitions from disorder to polar ordered waves and from waves to a polar ordered homogeneous state are marked by the red and magenta line, respectively. Parameters and simulation time for each point as in Fig. 1.

We also investigate the strength of the correlation parameters. As an example we present the two-particle correlation parameter C_2 as a function of density and noise strength in Fig. 3 and as a function of noise strength for fixed density in Fig. 4(a). Surprisingly, C_2 is not a monotonic function of the noise strength for fixed particle density. Instead it increases for decreasing noise strength until it reaches a maximum, goes through a minimum and then increases again. This behavior can be understood by studying the phase transitions of the Vicsek model. It is well known that there is a transition from disordered motion at high noise to polar ordered collective motion at smaller noise. The transition occurs discontinuously, has strong finite size effects and is closely related to the appearance of steep wave fronts [14,40]. It is absolutely plausible that these wave fronts go along with high correlations of many particles.

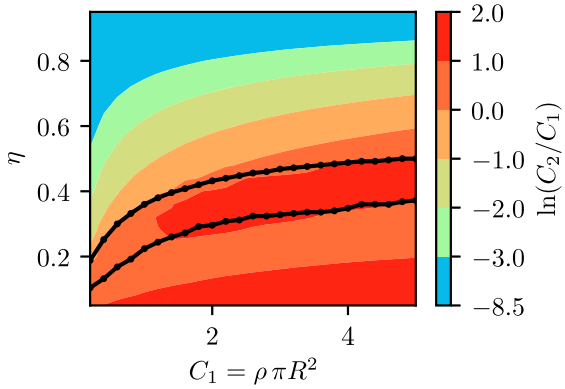

FIG. 3. Logarithm of the ratio between two-particle correlation parameter C_2 and mean number of neighbors C_1 for the standard Vicsek model. The solid black lines are the same transition lines as in Fig. 2. Data are from the same simulations.

A second transition from the wavelike pattern to the formation of homogeneously distributed clusters, also called polar liquid or the Toner-Tu phase, occurs for even smaller noise strength [13–15]. This second transition goes along with a drastic drop of correlations, cf. Fig. 4(a) or Fig. 3. Thus we understand the remarkable high correlation island between the two transition lines in Fig. 3. Consequently, it is possible to use the correlation parameter C_2 as a single order parameter to study both transitions. In fact, we find two local minima in the Binder cumulant of the two-particle correlation parameter C_2, indicating the location of both transitions, cf. Fig. 4(c). Remarkably, this second transition is not at all observed when only the polar order parameter or its Binder cumulant is studied; see Figs. 4(b) and 4(d).

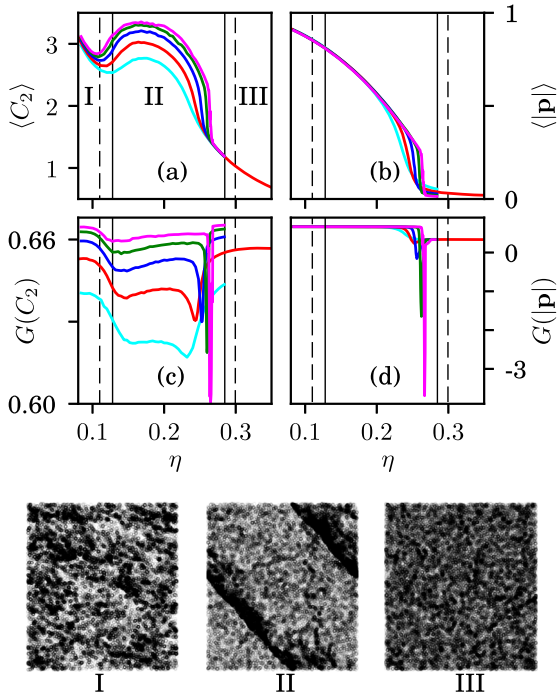

FIG. 4. Left: average value (a) and Binder cumulant (c) of the two-particle correlation parameter C_2. Right: average value (b) and Binder cumulant (d) of the polar order parameter. Bottom: Snapshots in the homogeneous polar ordered (I), the polar ordered wave front phase (II) and the disordered (III) phase. The vertical dashed line between phases I and II represents the transition point from [15] for $N = 10^4$ and the dashed line between II and III is the infinite system size transition of Ref. [15]. The solid vertical lines are the corresponding transitions obtained as the minima of the Binder cumulant of C_2 for the largest system $N = 8 \times 10^4$ (between I and II) and as an extrapolation to infinite system size (between II and III). Parameters: free path $v\tau = 0.5$, density $C_1 = \pi\rho = \pi/4$, particle number $N = 5 \times 10^3$ (cyan), $N = 10^4$ (red), $N = 2 \times 10^4$ (blue), $N = 4 \times 10^4$ (green), and $N = 8 \times 10^4$ (magenta). Average over 24 realizations, each started from random initial conditions, thermalized 10^5 and recorded for 10^6 time steps.

In Fig. 4 we chose a parameter set that was also studied in Ref. [15] for quantitative comparisons [41]. For other parameter sets, the two minima in the Binder cumulant are even much clearer, see Supplemental Material [27]. The dashed horizontal lines of Fig. 4 represent the transition points of Ref. [15] whereas we obtained the solid vertical lines by studying the two-particle correlation order parameter. The transition between polar liquid and wave-front phase was obtained in Ref. [15] as the end point of a hysteresis loop for a finite system ($N = 10\,000$) which, however, depends on the details with which the hysteresis loop is recorded. It is therefore a different definition of the transition point than the minimum in the Binder cumulant that we used to obtain a slightly different value that is well defined and independent on the measuring procedure. The polar ordering transition was obtained in Ref. [15] for infinitely large systems ($\eta_c = 0.299\,44$) [41]. We obtain a value of $\eta_c = 0.285$ by finite size scaling, however, more system sizes might be necessary to obtain an accurate extrapolation [27].

In summary, we exactly calculate the neighbor distribution of a homogeneous many-particle system in the thermodynamic limit by including correlations of finite but arbitrary order and hence *generalize* the Poisson distribution. Explicit formulas that incorporate up to k particle correlations are given. The novel distributions depend only on a few correlation parameters. We explicitly demonstrate how to sample these parameters. The analysis is applied to the Vicsek model of self-propelled particles. We create a quantitative correlation map of the model marking the minimal order of correlations required for each point in parameter space. Even in a large fraction of the disordered phase multiparticle correlations are important. We find furthermore, that correlation parameters serve as order parameters to accurately investigate both phase transitions of the model. We propose the use of the presented correlation analysis as a general technique in the study of phase transitions in and out of equilibrium and furthermore in the wide field of spatial point pattern analysis. Another possible application lies in the development of kinetic theories that describe low order correlations exactly and incorporate higher order ones by closure relations. Such closure approaches can be tested with the help of the measured correlation parameters.

The authors gratefully acknowledge the GWK support for funding this project by providing computing time through the Center for Information Services and HPC (ZIH) at TU Dresden on the HRSK-II. The authors gratefully acknowledge the Universitätsrechenzentrum Greifswald and the High Performance Computing facilities at C3-UNAM for providing computing time. M. Z. acknowledges Francisco Sevilla and Maximino Aldana for valuable discussions and CONACyT scholarship 592409/307643. We thank Michael Himpel for valuable discussions. We thank Alexandre Solon for valuable discussions and providing data of Ref. [15].

*Corresponding author.
ruediger.kuersten@uni-greifswald.de

[1] P. A. Egelstaff, Annu. Rev. Phys. Chem. **24**, 159 (1973).
[2] H. J. Bussemaker and M. H. Ernst, Phys. Rev. E **53**, 5837 (1996).
[3] T. P. C. Van Noije, M. H. Ernst, and R. Brito, Physica (Amsterdam) **251A**, 266 (1998).
[4] Y.-L. Chou and T. Ihle, Phys. Rev. E **91**, 022103 (2015).
[5] A. R. Denton and N. W. Ashcroft, Phys. Rev. A **39**, 426 (1989).
[6] J. Blawzdziewicz, B. Cichocki, and G. Szamel, J. Chem. Phys. **91**, 7467 (1989).
[7] S. Zhou and E. Ruckenstein, Phys. Rev. E **61**, 2704 (2000).
[8] K. Zahn, G. Maret, C. Ruß, and H. H. von Grünberg, Phys. Rev. Lett. **91**, 115502 (2003).
[9] A. Härtel, D. Richard, and T. Speck, Phys. Rev. E **97**, 012606 (2018).
[10] K. R. Mecke, T. Buchert, and H. Wagner, Astron. Astrophys. **288**, 697 (1994).
[11] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet, Phys. Rev. Lett. **75**, 1226 (1995).
[12] A. Czirók, H. E. Stanley, and T. Vicsek, J. Phys. A **30**, 1375 (1997).
[13] G. Grégoire and H. Chaté, Phys. Rev. Lett. **92**, 025702 (2004).
[14] H. Chaté, F. Ginelli, G. Grégoire, and F. Raynaud, Phys. Rev. E **77**, 046113 (2008).
[15] A. P. Solon, H. Chaté, and J. Tailleur, Phys. Rev. Lett. **114**, 068101 (2015).
[16] A. P. Solon, J.-B. Caussin, D. Bartolo, H. Chaté, and J. Tailleur, Phys. Rev. E **92**, 062111 (2015).
[17] P. J. Diggle, *Statistical Analysis of Spatial Point Patterns* (Academic Press, London, 1983).
[18] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan, *Statistical Analysis and Modelling of Spatial Point Patterns* (John Wiley & Sons, Chichester, 2008), Vol. 70.
[19] A. Baddeley, E. Rubak, and R. Turner, *Spatial Point Patterns: Methodology and Applications with R* (Chapman and Hall/CRC, Boca Raton, 2015).
[20] R. Law, J. Illian, D. F. Burslem, G. Gratzer, C. Gunatilleke, and I. Gunatilleke, J. Ecol. **97**, 616 (2009).
[21] E. Velázquez, I. Martínez, S. Getzin, K. A. Moloney, and T. Wiegand, Ecography **39**, 1042 (2016).
[22] P. J. Diggle, N. Lange, and F. M. Beneš, J. Am. Stat. Assoc. **86**, 618 (1991).
[23] M. Kerscher, J. Schmalzing, J. Retzlaff, S. Borgani, T. Buchert, S. Gottlöber, V. Müller, M. Plionis, and H. Wagner, Mon. Not. R. Astron. Soc. **284**, 73 (1997).
[24] J. Parker, E. Sherman, M. van der Raa, D. van der Meer, L. E. Samelson, and W. Losert, Phys. Rev. E **88**, 022720 (2013).
[25] H. Mantz, K. Jacobs, and K. Mecke, J. Stat. Mech. (2008) P12015.
[26] S. D. Poisson, *Probabilité des Jugements en Matière Criminelle et en Matière Civile, Précédées des Règles Générales du Calcul des Probabilitiés* (Bachelier, Paris, France, 1837).
[27] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.124.088002 for a derivation of Eqs. (3) and (4), sampling algorithms for the coefficients C_k and D_k and additional numerical results.

[28] B. R. A. Nijboer and L. Van Hove, Phys. Rev. **85**, 777 (1952).
[29] F. H. Ree and W. G. Hoover, J. Chem. Phys. **40**, 939 (1964).
[30] R. J. Wheatley, Phys. Rev. Lett. **110**, 200601 (2013).
[31] H. D. Ursell, Math. Proc. Cambridge Philos. Soc. **23**, 685 (1927).
[32] J. E. Mayer and E. Montroll, J. Chem. Phys. **9**, 2 (1941).
[33] R. Kürsten, arXiv:1910.13468.
[34] The overlap area of three arbitrarily placed circles generalizes the area formula of the famous Reuleaux triangle [35].
[35] F. Reuleaux, *Lehrbuch der Kinematik: Theoretische Kinematik: Grundzüge einer Theorie des Maschinenwesens. 1* (Vieweg, Braunschweig, 1875), Vol. 1.
[36] A. Bricard, J.-B. Caussin, N. Desreumaux, O. Dauchot, and D. Bartolo, Nature (London) **503**, 95 (2013).
[37] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).
[38] J. Toner and Y. Tu, Phys. Rev. Lett. **75**, 4326 (1995).
[39] J. Toner and Y. Tu, Phys. Rev. E **58**, 4828 (1998).
[40] T. Ihle, Phys. Rev. E **88**, 040303(R) (2013).
[41] In Ref. [15] the density was varied at constant noise strength. Therefore both transitions occur not exactly at the same density $\rho = 0.25$ that was used in Fig. 4. The polar ordering transition occurred at $\rho = 0.251$ for $\eta = 0.3$ in Ref. [15] from which we interpolated together with the next point $\eta_c = 0.29944$ for $\rho = 0.25$.