

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

Rüdiger Kürsten, Sven Stroteich, and Thomas Ihle  
*Institut für Physik, Universität Greifswald, Felix-Hausdorff-Str. 6, 17489 Greifswald, Germany*

Martín Zumaya Hernández  
*Instituto de Ciencias Físicas, Universidad Nacional Autónoma de México,  
Apartado Postal 48-3, Código Postal 62251, Cuernavaca, Morelos, México and  
Centro de Ciencias de la Complejidad, Universidad Nacional Autónoma de México, Ciudad de México, México*

(Dated: January 30, 2020)

## I. STRUCTURE OF THE SUPPLEMENTAL

This supplemental material is organized as follows. In Sec. II we give the details of the derivation of the main theoretical result, Eqs.(3) and (4). In Sec. III we provide efficient algorithms to sample the parameters of the distribution given by Eqs. (2) and (5). In Sec. IV we derive explicit formulas to calculate the overlap area of two and three disks in two dimensions. In Sec. V we precisely define the models that are investigated in this paper and present additional numerical results.

## II. SPATIAL CORRELATIONS

We assume that we have  $N$  identical particles that are in a random state. Furthermore we assume that  $N$  is large such that is reasonable to consider the limit  $N \rightarrow \infty$  when necessary. We denote the  $N$ -particle probability density by  $P_N(1, 2, \dots, N)$ , where 1 represents all degrees of freedom of particle one like e.g. position, velocity or orientation. Similarly  $2, \dots, N$  represent all degrees of freedom of particles 2 to  $N$ . We assume that  $P_N$  is invariant under permutations of its arguments which should be the case at least in a stationary state. The  $k$ -particle probability density is obtained from  $P_N$  by

$$P_k(1, 2, \dots, k) := \int P_N(1, 2, \dots, N) d(k+1) d(k+2) \dots dN, \quad (\text{S.1})$$

where the integration is performed over all possible configurations of the  $N-k$  particles. We assume furthermore that the one-particle probability density  $P_1$  is spatially homogeneous, that is it does not depend on the position of the considered particle. As long as there are no deterministic external forces and no walls all those assumption must be satisfied for finite ergodic systems in the stationary state.

We define the  $k$ -particle correlation function  $G_k$  by

$$\begin{aligned} G_1(1) &:= P_1(1), \\ G_k(1, \dots, k) &:= P_k(1, \dots, k) \\ &\quad - \text{all possible combinations of} \\ &\quad \{G_1, \dots, G_{k-1}\}, \\ &\quad \text{where each of the arguments } \{1, \dots, k\} \\ &\quad \text{appears exactly once.} \end{aligned} \quad (\text{S.2})$$

For example we have for two-particle and three-particle correlation function

$$\begin{aligned} G_2(1, 2) &:= P_2(1, 2) - P_1(1)P_1(2), \\ G_3(1, 2, 3) &:= P_3(1, 2, 3) - P_1(1)G_2(2, 3) - P_1(2)G_2(1, 3) \\ &\quad - P_1(3)G_2(1, 2) - P_1(1)P_1(2)P_1(3), \end{aligned} \quad (\text{S.3})$$

which is also called cluster-expansion. For a homogeneous system, we can express  $G_2$  depending only on the positions  $\mathbf{r}_1$  and  $\mathbf{r}_2$  in terms of the pair correlation function also called radial distribution function  $g(r)$

$$G_2(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{V^2} [g(|\mathbf{r}_2 - \mathbf{r}_1|) - 1]. \quad (\text{S.4})$$

In equilibrium thermodynamics, the knowledge of  $g(r)$  is sufficient to obtain the equation of state as long as only two-particle interactions are involved. The calculation of certain quantities such as thermal expansion coefficient or heat capacity involves partial derivatives of the pair correlation function with respect to pressure or temperature. Those are directly related to expressions involving higher order correlation functions. Thus, one might argue that certain macroscopic quantities depend directly on higher order correlation functions even in equilibrium. On the other hand, knowing the pair correlation function for all pressures and temperatures is equivalent in equilibrium. Hence, strictly speaking, it is not necessary to study higher order correlations of systems with two-particle interactions. From a practical point of view, it might be reasonable to study higher order correlations any way, if the pair correlation function is known only for special values of temperature and pressure or if multi-particle correlations are involved.

Reformulating Eq. (S.2) one can also obtain the  $k$ -particle probability from the one-particle probability and the correlation functions. For example one finds the recursion relation

$$P_k(1, \dots, k) = \sum_{l=1}^k G_l(1, \dots, l) P_{k-l}(l+1, \dots, k) + \text{Permutations}, \quad (\text{S.5})$$

where we sum over permutations of  $\{2, \dots, k\}$  that are not only interchanging the arguments of each function, but exchange at least one argument of  $G_l$  and  $P_{k-l}$ .

From the above definitions it is straight forward to find

$$\begin{aligned} \int G_2(1, 2) d1 &= P_1(2) - P_1(2) = 0 \\ &= P_1(1) - P_1(1) = \int G_2(1, 2) d2. \end{aligned} \quad (\text{S.6})$$

Similarly, one can show by induction that in general

$$\begin{aligned} \int G_k(1, \dots, k) d1 &= \int G_k(1, \dots, k) d2 = \dots \\ &= \int G_k(1, \dots, k) dN = 0. \end{aligned} \quad (\text{S.7})$$

The mean field assumption

$$P_N(1, 2, \dots, N) = P_1(1)P_1(2) \dots P_1(N) \quad (\text{S.8})$$

is identical to setting all correlation functions to zero

$$G_2 = G_3 = \dots = G_N = 0. \quad (\text{S.9})$$

Considering the full hierarchy of correlation functions is extremely complex and the question arises which of the correlation functions are important and have to be considered. Is mean field already an excellent approximation? Or is it necessary to consider at least  $G_2$ ? Or also  $G_3$ ? Or do we need to take all correlations into account to obtain a good description of the system?

The correlation functions  $G_2, G_3, \dots, G_N$  are high dimensional and therefore hard to measure. Furthermore, it is not clear if it is important or not for the dynamics of  $P_1$  if one of the correlation functions is large at some particular values of its arguments.

However, the microscopic dynamics of a particle depends explicitly on the number of its neighbors. Hence, we know for sure that the probability distribution of the number of neighbors has an explicit influence on the dynamics of  $P_1$ . Therefore we study the influence of the correlation functions  $G_2, G_3, \dots$  on the number of neighbor probability distribution.

### A. Bernoulli experiment

The mean field assumption (S.9) corresponds to the random placement of independent particles. Asking for

the number of particles in a circle of radius  $R$  we will find the answer to be Poisson-distributed in the limit  $N \rightarrow \infty$ . In the mean field case the number of neighbors of a given particle have the same probability distribution.

Here, however, we take correlations into account. As a first step we consider a nonzero  $G_2$ , however, we assume that higher correlations are zero, that is  $G_3 = G_4 = \dots = G_N = 0$ . In this case, we can still calculate the distribution of particles within a circle exactly in the limit of  $N \rightarrow \infty$ . It is useful to define the parameter  $C_k$  according to

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

In particular we get for the two-particle correlations

$$C_2 := N^2 \int G_2(1, 2) d1 d2 \theta(R - |\mathbf{r}_1|) \theta(R - |\mathbf{r}_2|), \quad (\text{S.10})$$

where  $\theta$  denotes the Heavyside function. The coefficient might also be expressed in terms of the pair correlation function for homogeneous systems as

$$C_2 = \left(\frac{N}{V}\right)^2 \int_{\mathbb{R}^2} [g(|\mathbf{r}_2 - \mathbf{r}_1|) - 1] \times \theta(R - |\mathbf{r}_1|) \theta(R - |\mathbf{r}_2|) d\mathbf{r}_1 d\mathbf{r}_2. \quad (\text{S.11})$$

It is worth noting that changing one of the spatial integration regions to the outside of the circle yields

$$\begin{aligned} -C_2 &= N^2 \int G_2(1, 2) d1 d2 \theta(R - |\mathbf{r}_1|) [1 - \theta(R - |\mathbf{r}_2|)] \\ &= N^2 \int G_2(1, 2) d1 d2 [1 - \theta(R - |\mathbf{r}_1|)] \theta(R - |\mathbf{r}_2|) \end{aligned} \quad (\text{S.12})$$

due to the property (S.7) of  $G_2$ . Integrating both spatial coordinates outside the circle yields again

$$C_2 = N^2 \int G_2(1, 2) d1 d2 [1 - \theta(R - |\mathbf{r}_1|)] \times [1 - \theta(R - |\mathbf{r}_2|)]. \quad (\text{S.13})$$

The probability to find  $s$  particles within a circle is given by

$$p(s) = \binom{N}{s} \int d1 \dots dN P_N(1, \dots, N) \theta_1 \dots \theta_s \times (1 - \theta_{s+1}) \dots (1 - \theta_N), \quad (\text{S.14})$$

where we introduced the abbreviation

$$\theta_k := \theta(R - |\mathbf{r}_k|) \quad (\text{S.15})$$

and we included the combinatorial factor  $\binom{N}{s}$  to take care of the fact that any selection of  $s$  particles can be

integrated over the inside of the circle and the rest over the outside. Inserting a sum over all possible combinations of  $P_1$  and  $G_2$  for  $P_N$ , that is setting  $G_3 = G_4 = \dots = G_N = 0$  in the full hierarchy of  $P_N$ , Eq. (S.2), we obtain

$$\begin{aligned}
p(s) &= \sum_{k_2=0}^{\infty} \left(\frac{C_2}{N^2}\right)^{k_2} \binom{N}{2k_2} \binom{2k_2}{k_2} \frac{k_2!}{2^{k_2}} \\
&\times \sum_{k_1=0}^{\infty} \left(-\frac{C_2}{N^2}\right)^{k_1} \binom{N-2k_2}{2k_1} \binom{2k_1}{k_1} k_1! \\
&\times \sum_{k_0=0}^{\infty} \left(\frac{C_2}{N^2}\right)^{k_0} \binom{N-2k_2-2k_1}{2k_0} \binom{2k_0}{k_0} \frac{k_0!}{2^{k_0}} \\
&\times \sum_{q_1=0}^{\infty} \left(\frac{C_1}{N}\right)^{q_1} \binom{N-2k_2-2k_1-2k_0}{q_1} \\
&\times \sum_{q_0=0}^{\infty} \left(1 - \frac{C_1}{N}\right)^{q_0} \delta(s-2k_2-k_1-q_1) \\
&\times \delta(N-2k_2-2k_1-2k_0-q_1-q_0) \quad (\text{S.16})
\end{aligned}$$

with the convention that  $\binom{N}{l} = 0$  if  $l > N$ .  $k_2$  represents the number of pairs ( $G_2$ ) where both arguments are integrated over the inside of the circle and the following combinatorial factor is the number of possibilities to choose  $k_2$  pairs from  $N$  particles.  $k_1$  is the number of pairs with one argument integrated over the circles inside and one argument integrated over the outside. The following combinatorial factor gives the number of possible choices of  $k_1$  ordered pairs from the remaining  $N-2k_2$  particles.  $k_0$  represents the number of pairs where both arguments are integrated over the outside of the circle, analogously  $q_1$  and  $q_0$  are the numbers of single particles ( $P_1$ ) integrated over the circles inside and outside, respectively. There, we also include the corresponding combinatorial factors. The first  $\delta$ -function takes care that there are exactly  $s$  particles inside and the second one, that there are exactly  $N$  particles in total. All the combinatorial factors can be combined to

$$\frac{N!}{(N-2k_0-2k_1-2k_2-q_1)!q_1!k_0!k_1!k_2!2^{k_0}2^{k_2}}. \quad (\text{S.17})$$

For large  $N$  we can approximate

$$\begin{aligned}
\frac{N!}{(N-2k_0-2k_1-2k_2-q_1)!} &= N \cdot (N-1) \cdot \dots \\
&\times (N-2k_0-2k_1-2k_2-q_1+1) \\
&\approx N^{2k_0+2k_1+2k_2+q_1}. \quad (\text{S.18})
\end{aligned}$$

All corrections are of lower order in  $N$  and are eventually going to zero as  $N \rightarrow \infty$ . Inserting (S.18) in (S.16) and

performing the sum over  $q_0$  with the second  $\delta$  we obtain

$$\begin{aligned}
p(s) &= \sum_{k_2=0}^{\infty} \left(\frac{C_2}{2}\right)^{k_2} \sum_{k_1=0}^{\infty} (-C_2)^{k_1} \sum_{k_0=0}^{\infty} \left(\frac{C_2}{2}\right)^{k_0} \sum_{q_1=0}^{\infty} C_1^{q_1} \\
&\times \left(1 - \frac{C_1}{N}\right)^{N-2k_0-2k_1-2k_2-q_1} \frac{1}{k_0!k_1!k_2!q_1!} \\
&\times \delta(s-2k_2-k_1-q_1). \quad (\text{S.19})
\end{aligned}$$

In the limit  $N \rightarrow \infty$  we obtain

$$\left(1 - \frac{C_1}{N}\right)^{N-2k_0-2k_1-2k_2-q_1} \rightarrow \exp(-C_1) \quad (\text{S.20})$$

such that we find

$$\begin{aligned}
p(s) &= \sum_{k_0=0}^{\infty} \left(\frac{C_2}{2}\right)^{k_0} \frac{\exp(-C_2/2)}{k_0!} \sum_{k_2=0}^{\infty} \left(\frac{C_2}{2}\right)^{k_2} \\
&\times \sum_{k_1=0}^{\infty} (-C_2)^{k_1} \sum_{q_1=0}^{\infty} C_1^{q_1} \exp(-C_1) \exp(C_2/2) \frac{1}{k_1!k_2!q_1!} \\
&\times \delta(s-2k_2-k_1-q_1). \quad (\text{S.21})
\end{aligned}$$

The sum over  $k_0$  is just a sum over a Poisson distribution which gives one. Performing also the sum over  $q_1$  with the  $\delta$  we obtain

$$\begin{aligned}
p(s) &= C_1^s \exp(-C_1) \exp(C_2/2) \sum_{k_2=0}^{\infty} \left(\frac{C_2}{2C_1^2}\right)^{k_2} \frac{1}{(k-2k_2)!} \\
&\times \frac{1}{k_2!} \sum_{k_1=0}^{k-2k_2} \left(-\frac{C_2}{C_1}\right)^{k_1} 1^{s-2k_2-k_1} \frac{(s-2k_2)!}{k_1!(s-2k_2-k_1)!}. \quad (\text{S.22})
\end{aligned}$$

Using the binomial theorem to perform the sum over  $k_1$  we obtain

$$\begin{aligned}
p(s) &= C_1^s \exp(-C_1) \exp(C_2/2) \sum_{k_2=0}^{\infty} \left(\frac{C_2}{2C_1^2}\right)^{k_2} \\
&\times \frac{1}{k_2!(s-2k_2)!} \left(1 - \frac{A}{M}\right)^{k-2k_2} \\
&= (C_1 - C_2)^s \exp(C_2/2 - C_1) \\
&\times \sum_{k_2=0}^{\infty} \left[\frac{C_2}{2(C_1 - C_2)^2}\right]^{k_2} \frac{1}{k_2!(s-2k_2)!} \quad (1)
\end{aligned}$$

Eventually, this sum can be expressed as

$$\begin{aligned}
p(s) &= \frac{C_1^s \exp(-C_1)}{s!} \exp(C_2/2) \left(1 - \frac{C_2}{C_1}\right)^s \\
&\times \left(-\frac{(C_2 - C_1)^2}{2C_2}\right)^{(1-s)/2} \\
&\times U\left[(1-k)/2, 3/2, -(C_2 - C_1)^2/(2C_2)\right], \quad (\text{S.23})
\end{aligned}$$

where  $U(\dots)$  is the confluent hypergeometric function of the second kind.

Considering the form of  $p(s)$  given by Eq. (1) it is possible to calculate the characteristic function  $\chi(u) := \sum_s p(s) \exp(ius)$  exactly by first performing the sum over  $s$  and than the remaining sum over  $k_2$ . The surprisingly simple result is given by

$$\begin{aligned} \chi(u) &= \langle \exp(ius) \rangle \\ &= \exp \left[ -C_1 + \frac{C_2}{2} + (C_1 - C_2) \exp(iu) \right. \\ &\quad \left. + \frac{C_2}{2} \exp(i2u) \right]. \end{aligned} \quad (\text{S.24})$$

Taking into account nonzero three-particle correlations, however, still assuming  $G_4 = G_5 = \dots = G_N = 0$  we find in complete analogy to Eq.(S.16) the probability of finding exactly  $s$  particles within a given circle of radius  $R$  as

$$\begin{aligned} p(s) &= \sum_{l_0=0}^{\infty} \left( \frac{-C_3}{N^3} \right)^{l_0} \binom{N}{3l_0} \binom{3l_0}{l_0} \binom{2l_0}{l_0} \frac{(l_0!)^2}{6^{l_0}} \\ &\times \sum_{l_1=0}^{\infty} \left( \frac{C_3}{N^3} \right)^{l_1} \binom{N-3l_0}{3l_1} \binom{3l_1}{l_1} \binom{2l_1}{l_1} \frac{(l_1!)^2}{2^{l_1}} \\ &\times \sum_{l_2=0}^{\infty} \left( \frac{-C_3}{N^3} \right)^{l_2} \binom{N-3l_0-3l_1}{3l_2} \binom{3l_2}{l_2} \binom{2l_2}{l_2} \frac{(l_2!)^2}{2^{l_2}} \\ &\times \sum_{l_3=0}^{\infty} \left( \frac{C_3}{N^3} \right)^{l_3} \binom{N-3l_0-3l_1-3l_2}{3l_3} \\ &\times \binom{3l_3}{l_3} \binom{2l_3}{l_3} \frac{(l_3!)^2}{6^{l_3}} \\ &\times \sum_{k_0=0}^{\infty} \left( \frac{C_2}{N^2} \right)^{k_0} \binom{N-3l}{2k_0} \binom{2k_0}{k_0} \frac{k_0!}{2^{k_0}} \\ &\times \sum_{k_1=0}^{\infty} \left( -\frac{C_2}{N^2} \right)^{k_1} \binom{N-3l-2k_0}{2k_1} \binom{2k_1}{k_1} k_1! \\ &\times \sum_{k_2=0}^{\infty} \left( \frac{C_2}{N^2} \right)^{k_2} \binom{N-3l-2k_0-2k_1}{2k_2} \binom{2k_2}{k_2} \frac{k_2!}{2^{k_2}} \\ &\times \sum_{q_0=0}^{\infty} \left( 1 - \frac{C_1}{N} \right)^{q_0} \binom{N-3l-2k}{q_0} \\ &\times \sum_{q_1=0}^{\infty} \left( \frac{C_1}{N} \right)^{q_1} \delta(s-l_1-2l_2-3l_3-k_1-2k_2-q_1) \\ &\times \delta(N-3l-2k-q), \end{aligned} \quad (\text{S.25})$$

where  $l = l_0 + l_1 + l_2 + l_3$ ,  $k = k_0 + k_1 + k_2$  and  $q = q_0 + q_1$ . Here  $l_i$  denotes the number of correlated triplets with exactly  $i$  particles inside the circle. Analogously  $k_i$  and  $q_j$  are the numbers of correlated pairs or singlets with exactly  $i$  and  $j$  particles inside the circle, respectively.

Simplifying the combinatorial factors we obtain

$$\begin{aligned} p(s) &= \sum_{\dots} \left( \frac{-C_3}{6} \right)^{l_0} \left( \frac{C_3}{2} \right)^{l_1} \left( \frac{-C_3}{2} \right)^{l_2} \left( \frac{C_3}{6} \right)^{l_3} \left( \frac{C_2}{2} \right)^{k_0} \\ &\times (-C_2)^{k_1} \left( \frac{C_2}{2} \right)^{k_2} \left( 1 - \frac{C_1}{N} \right)^{q_0} C_1^{q_1} \\ &\times \frac{N!}{(N-3l-2k-q_0)! N^{3l+2k+q_1}} \\ &\times \frac{\delta(s-l_1-2l_2-3l_3-k_1-2k_2-q_1)}{l_0! l_1! l_2! l_3! k_0! k_1! k_2! q_0!} \\ &\times \delta(N-3l-2k-q). \end{aligned} \quad (\text{S.26})$$

Performing the sum over  $q_0$  with the second  $\delta$  we obtain

$$\begin{aligned} p(s) &= \sum_{\dots} \left( \frac{-C_3}{6} \right)^{l_0} \left( \frac{C_3}{2} \right)^{l_1} \left( \frac{-C_3}{2} \right)^{l_2} \left( \frac{C_3}{6} \right)^{l_3} \left( \frac{C_2}{2} \right)^{k_0} \\ &\times (-C_2)^{k_1} \left( \frac{C_2}{2} \right)^{k_2} \left( 1 - \frac{C_1}{N} \right)^{N-3l-2k-q_1} C_1^{q_1} \\ &\times \frac{N!}{(N-3l-2k-q_1)! N^{3l+2k+q_1}} \\ &\times \frac{\delta(s-l_1-2l_2-3l_3-k_1-2k_2-q_1)}{l_0! l_1! l_2! l_3! k_0! k_1! k_2! q_1!}. \end{aligned} \quad (\text{S.27})$$

Performing the limit  $N \rightarrow \infty$  we obtain

$$\begin{aligned} \left( 1 - \frac{C_1}{N} \right)^{N-3l-2k-q_1} &\rightarrow \exp(-C_1) \\ \frac{N!}{(N-3l-2k-q_1)! N^{3l+2k+q_1}} &\rightarrow 1. \end{aligned} \quad (\text{S.28})$$

Performing the sums over  $l_0$  and  $k_0$

$$\begin{aligned} \sum_{l_0=0}^{\infty} \frac{(-C_3/6)^{l_0}}{l_0!} &= \exp(-C_3/6), \\ \sum_{k_0=0}^{\infty} \frac{(C_2/2)^{k_0}}{k_0!} &= \exp(C_2/2) \end{aligned} \quad (\text{S.29})$$

we obtain

$$\begin{aligned} p(s) &= \sum_{\dots} \left( \frac{C_3}{2} \right)^{l_1} \left( \frac{-C_3}{2} \right)^{l_2} \left( \frac{C_3}{6} \right)^{l_3} (-C_2)^{k_1} \\ &\times \left( \frac{A}{2} \right)^{k_2} C_1^{q_1} \exp(-C_1 + C_2/2 - C_3/6) \\ &\times \frac{\delta(s-l_1-2l_2-3l_3-k_1-2k_2-q_1)}{l_1! l_2! l_3! k_1! k_2! q_1!}. \end{aligned} \quad (\text{S.30})$$

Performing the sum over  $q_1$  with the  $\delta$  and the binomial

sums over  $l_1$  and  $k_1$  we find

$$\begin{aligned}
p(s) &= \sum_{l_2 l_3 k_2} \exp(-C_1 + C_2/2 - C_3/6) \left[ C_1 + \frac{C_3}{2} - C_2 \right]^s \\
&\times \left[ -\frac{2C_3}{(2C_1 + C_3 - 2C_2)^2} \right]^{l_2} \left[ \frac{4C_3}{3} \frac{1}{(2C_1 + C_3 - 2C_2)^3} \right]^{l_3} \\
&\times \left[ \frac{2C_2}{(2C_1 + C_3 - 2C_2)^2} \right]^{k_2} \\
&\times \frac{1}{l_2! l_3! k_2! (s - 2l_2 - 3l_3 - 2k_2)!}. \tag{S.31}
\end{aligned}$$

It is possible to represent one more sum in terms of a confluent hypergeometric function, however, we can also handle the three remaining sums numerically. Furthermore, although Eq. (S.31) looks much more complicated than Eq. (1) it is still possible to calculate the characteristic function exactly by first performing the sum over  $s$  and then all the remaining sums over  $k_2, l_2, l_3$  resulting in

$$\begin{aligned}
\chi(u) &= \exp \left[ -C_1 + \frac{C_2}{2} - \frac{C_3}{6} + \left( C_1 - C_2 + \frac{C_3}{2} \right) e^{iu} \right. \\
&\quad \left. + \left( \frac{C_2}{2} - \frac{C_3}{2} \right) e^{i2u} + \frac{C_3}{6} e^{i3u} \right]. \tag{S.32}
\end{aligned}$$

Including additionally correlations of order  $k$  the procedure is completely analogous. There appear  $k+1$  more sums in the analog of Eq. (S.25). Two of them can be calculated immediately, such that one remains with  $k-1$  additional sums. However, calculating the characteristic function one can perform all this sums exactly and it is possible to show by induction that

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

Eventually, the distribution of the number of particles found in a ball of Radius  $R$  is obtained by transforming back the characteristic function

$$p(s) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{u=0}^{n-1} \exp(-ius) \chi(u). \tag{S.33}$$

## B. Number of neighbor distribution

If the particles were independent the distribution (S.23) would already give the neighbor distribution of a given particle. With nonzero pair correlations, however, the neighbor distribution differs. The neighborhood dis-

tribution is given by

$$\begin{aligned}
p_n(k) &= \langle \delta(k - \sum_{j=2}^N \theta_{1j}) \rangle \\
&= \int P_N(1, 2, \dots) \delta(k - \sum_{j=2}^N \theta_{1j}) \\
&= \int P_1(1) P_{N-1}(2, 3, \dots) \delta(k - \sum_{j=2}^N \theta_{1j}) \\
&\quad + \int \sum_{l=2}^N G_2(1, l) P_{N-2}(2, \dots, l-1, l+1, \dots) \\
&\quad \times \delta(k - \sum_{j=2}^N \theta_{1j}) \\
&= \int P_1(1) P_{N-1}(2, 3, \dots) \delta(k - \sum_{j=2}^N \theta_{1j}) \\
&\quad + \int (N-1) G_2(1, 2) P_{N-2}(3, 4, \dots) \delta(k - \sum_{j=2}^N \theta_{1j}), \tag{S.34}
\end{aligned}$$

where integration is performed over all coordinates. Similar to Eq. (S.15) we introduced the abbreviation

$$\theta_{jk} := \theta(R - |\mathbf{r}_k - \mathbf{r}_j|). \tag{S.35}$$

In the limit  $N \rightarrow \infty$  we can replace  $N-1$  by  $N$ . The first term in the above expression corresponds to distribution of an Bernoulli experiment, however, the second term gives a correction

$$\begin{aligned}
p_n(k) &= p(k) \\
&\quad + \int N G_2(1, 2) \theta_{12} P_{N-2}(3, 4, \dots) \delta(k - \sum_{j=2}^N \theta_{1j}) \\
&\quad + \int N G_2(1, 2) (1 - \theta_{12}) P_{N-2}(3, 4, \dots) \delta(k - \sum_{j=2}^N \theta_{1j}) \\
&= p(k) \\
&\quad + \int N G_2(1, 2) \theta_{12} P_{N-2}(3, 4, \dots) \delta(k - 1 - \sum_{j=3}^N \theta_{1j}) \\
&\quad + \int N G_2(1, 2) (1 - \theta_{12}) P_{N-2}(3, 4, \dots) \delta(k - \sum_{j=3}^N \theta_{1j}) \\
&= p(k) + \int N G_2(1, 2) \theta_{12} P_{N-2}(3, 4, \dots) \\
&\quad \times [\delta(k - 1 - \sum_{j=3}^N \theta_{1j}) - \delta(k - \sum_{j=3}^N \theta_{1j})], \\
&= p(k) + [p(k-1) - p(k)] N \int G_2(1, 2) \theta_{12}, \tag{S.36}
\end{aligned}$$

where it should be noted that  $p(k) = 0$  for  $k < 0$ .

For generalizing this result taking into account correlations up to order  $l_{max}$  we introduce the abbreviations

$$D_k := N^{k-1} \int G_k(1, 2, \dots, k) \theta_{12} \dots \theta_{1k} d1 \dots dk, \quad (5)$$

where  $D_1 = 1$  and  $D_2$  can be expressed in terms of the pair correlation function for homogeneous systems as

$$D_2 = 2\pi \left( \frac{N}{V} \right)^2 \int_0^R r [g(r) - 1] dr. \quad (S.37)$$

In analogy to Eqs. (S.34) and (S.36) we find in this case

$$p_n(k) = \sum_{l=1}^{l_{max}} \sum_{t=0}^{l-1} \binom{l-1}{t} (-1)^{t+l+1} \frac{1}{(l-1)!} p(k-t) D_l. \quad (S.38)$$

It is straight forward to calculate the characteristic function of the number of neighbor distribution

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

where  $\chi(u)$  is given by Eq. (3). Hence we expressed the number of neighbor distribution in terms of the coefficients  $C_l$  and  $D_l$ . The expression is exact in the limit  $N \rightarrow \infty$  if the higher order correlation functions  $G_l \equiv 0$  vanish for  $l > l_{max}$ .

### C. Minkowski Functionals

We characterized the correlations of a set of points on a given length scale by a view parameters,  $C_k$  and  $D_k$ . In this subsection we mention an alternative approach that is used to classify spatial point patterns. In particular for applications in cosmology, Minkowski functionals are widely used to characterize spatial point patterns, see e.g. Ref. [10] of the letter or [S1]. The idea is to draw a ball of radius  $r$  around each point. The union of all those balls is a subset of  $\mathbb{R}^3$  in three dimensions. In that case, there exist four Minkowski functionals, that can be used to characterize this subset: volume, surface, integrated mean curvature and Euler characteristics. If the Minkowski functionals are evaluated for different values of  $r$  they contain information on the point set and in particular also on their correlations. However, to obtain the full information one would need to evaluate the functionals for all radii. An advantage of the method presented here is that the coefficients  $C_k, D_k$  contain correlation information on a given length scale (that might be known to be relevant due to the microscopic dynamics of a system) and on the (arbitrary) desired order.

### D. Applicability for long range interactions

We defined the correlation parameters  $C_k, D_k$  on a length scale  $R$ . This is motivated by the fact that in many situations the microscopic interactions possess an intrinsic length scale that is important for the dynamics of the system. However, if the interactions are long range (like e.g. Coulomb interactions) one might still apply the analysis technique. Long range interactions do not automatically imply that there is no characteristic length scale. Such a length scale might be selected by a complex behavior of the system. If one has identified a characteristic length scale for a concrete given system one can perform the correlation analysis on that length scale. An alternative approach is to compare the results of the correlation analysis for many different length scales.

### III. SAMPLING PARAMETERS $C_l$ AND $D_l$

It is not possible to sample the parameters  $C_l$  and  $D_l$  directly. However, they are related to two types of quantities that can be sampled with high efficiency. The first quantities are expectation values of polynomials of the number of neighbors

$$\mu_l := \left\langle \frac{k!}{(k-l)!} \right\rangle = \sum_{k=0}^{N-1} p_n(k) \frac{k!}{(k-l)!} \quad (S.39)$$

for  $l = 1, 2, \dots$ . It is straight forward to sample these quantities by just counting the number of neighbors for each particle in each time step and averaging the corresponding polynomial of this number.

The second quantities are sums of the overlap volume of balls of radius  $R$  drawn around all ordered  $l$ -plets of different particles divided by the systems Volume

$$V_o(l) := \frac{1}{V} \left\langle \sum_{k_1, \dots, k_l, k_i \neq k_j} \mathcal{V}_{\text{overlap}}(k_1, \dots, k_l) \right\rangle, \quad (S.40)$$

where  $\mathcal{V}_{\text{overlap}}(k_1, \dots, k_l)$  denotes the volume of the overlap of  $l$  balls of radius  $R$  centered at the positions of the particles  $k_1, \dots, k_l$ . There are two efficient ways to sample these quantities. One can either take all  $l$ -plets and calculate the overlap volume exactly. For example for pairs, the overlap only depends on the distance between the particles and for triplets it depends only on the three distances between the three particles. In two dimensions we derive explicit formulas for the two and three particle overlap areas in Sec. IV.

There is an alternative Monte Carlo approach that is in particular suitable to sample also  $V_o(l)$  for large  $l$ . Assume we chose an ordered triplet of particles, e.g.  $(7, 3, 5)$  and want to estimate  $\frac{\mathcal{V}_{\text{overlap}}(7, 3, 5)}{V}$ . We can place a virtual particle at a random position and count one if

	$V_o(2)$
Direct Measurement	7.67(17)
Monte Carlo Sampling	7.50(28)

TABLE I. Comparison of the direct measurement and the Monte Carlo sampling of  $V_o(2)$  for the standard Vicsek model. We give the standard deviation of the mean value in brackets. Both approaches gives similar results, except the standard deviation of the Monte Carlo sampling is approximately 50% larger. System parameters:  $N = 22500$ ,  $\eta = 0.5$ ,  $v\tau/R = 5$ ,  $C_1 = 5$ , thermalization for  $10^5$  and measurement for  $10^6$  time steps for 25 realizations.

the virtual particle lies in the overlap of  $(7, 3, 5)$  that is if it has a distance smaller than  $R$  to each of the particles 7, 3, 5. Otherwise we count zero. If we repeat the procedure  $n$  times and divide the sum by  $n$  we obtain an estimate of the above mentioned volume fraction.

In order to obtain an estimate for the sum in Eq. (S.40), in principle, we need to repeat the procedure for each ordered triplet. We can instead just sum over all unordered triplets and multiply by  $l!$ , where  $l = 3$  in the case of triplets. Furthermore, we reach an extreme improvement in performance using the same randomly placed virtual particles for all triplets. In that way, we just need to find out in the overlap of how many triplets of real particles does a randomly placed real particle lie? This number depends only on the number of real neighbors of the virtual particle. Assume the virtual particle has  $m$  real neighbors, than it lies in the overlap of  $\binom{m}{l}$  unordered or  $\binom{m}{l}l!$  ordered triplets of real particles, where  $l = 3$  for triplets. In general we find

$$V_o(l) = \left\langle \binom{m}{l} l! \right\rangle, \quad (\text{S.41})$$

where  $m$  is the number of real neighbors of a randomly placed virtual particle and the expectation value is taken over many virtual particles and a time series of the positions of the real particles. Usually it does not significantly slow down molecular dynamics simulations if we use as many virtual particles as we have real particles. In Table I we show an example of a measurement of  $V_o(2)$ . We compare the values obtained by directly measuring the overlap areas with the Monte Carlo approach. The measured values agree within error bars. As expected, the uncertainty in the Monte Carlo sampling is a bit larger (by a factor of 1.5) but on the same order of magnitude. On the other hand, the Monte Carlo approach has the advantage that one can sample  $V_o(l)$  for all  $l$  at the same time.

We can directly connect the correlation coefficients  $C_l$  to the sampled overlap areas. We rewrite their definition

as

$$\begin{aligned} C_l &= N^l \int G_l \theta_1 \dots \theta_l d1 \dots dl \\ &= N^l \int G_l \theta_{01} \dots \theta_{0l} d1 \dots dl \\ &= \frac{N^l}{V} \int G_l \theta_{01} \dots \theta_{0l} d0 d1 \dots dl, \end{aligned} \quad (\text{S.42})$$

where we invented a virtual particle 0 in the second line. Due to translational invariance the integral does not depend on the position of this virtual particle. Hence we can also integrate the virtual particles position over the whole space and compensate for it by dividing by the total volume. For large  $N$  we have furthermore

$$\begin{aligned} &\frac{N^l}{V} \int P_l \theta_{01} \dots \theta_{0l} d0 d1 \dots dl \\ &= \frac{N^l}{V} \int P_N \theta_{01} \dots \theta_{0l} d0 d1 \dots dN \\ &\approx V_o(l), \end{aligned} \quad (\text{S.43})$$

since the integral divided by  $V$  gives the probability that a virtual particle is in the overlap of the particles  $(1, \dots, l)$ . However, for large  $N$  there are approximately  $N^l$  possibilities to choose an ordered  $l$ -plet from  $N$  particles.

Inserting the expansion (S.5) for  $G_l$  into Eq. (S.42) and using Eq. (S.43) we obtain

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

where  $\binom{l-1}{k-1}$  gives the number of permutations in (S.5) and  $V_o(1) := B_d(R) \frac{N}{V}$  with  $B_d(R)$  being the volume of a ball of radius  $R$  in  $d$  dimensions.

Inserting the expansion (S.5) into the definition of the coefficients  $D_l$ , Eq. (5), we obtain analogously

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

where

$$\begin{aligned} \mu_{l-1} &= N^{l-1} \int P_l \theta_{12} \dots \theta_{1l} d1 \dots dl \\ &= N^{l-1} \int P_N \theta_{12} \dots \theta_{1l} d1 \dots dN \end{aligned} \quad (\text{S.44})$$

is  $N^{l-1}$  times the probability that particles  $(2, \dots, l)$  are neighbors of particle one. However this can be rewritten

as

$$\begin{aligned}
\mu_{l-1} &= N^{l-1} \sum_s p_n(s) \frac{1}{\binom{N-1}{s}} \binom{N-1-(l-1)}{s-(l-1)} \\
&= N^{l-1} \sum_s p_n(s) \frac{s!(N-1-s)!}{(N-1)!} \\
&\quad \times \frac{(N-l)!}{(N-s-1)!(s-l+1)!} \\
&= \sum_s p_n(s) \frac{s!}{(s-l+1)!} \tag{S.45}
\end{aligned}$$

which coincides with the previous definition of  $\mu_l$ , Eq. (S.39). In the first line we have the probability that there are  $s$  neighbors of particle one times the probability that these neighbors are exactly  $2, 3, \dots, l, n_1, n_2, \dots, n_{s-(l-1)}$  times the number of possible choices for  $n_1, n_2, \dots, n_{s-(l-1)}$ .

Eqs. (6) and (7) are a system of linear equations relating the number of neighbor distribution parameters  $C_l$  and  $D_l$  to the measured quantities  $V_o(l)$  and  $\mu_l$ . We give the coefficients explicitly up to  $l = 5$  as

$$\begin{aligned}
C_1 &= B_d(R) \frac{N}{V}, \\
C_2 &= V_o(2) - C_1^2, \\
C_3 &= V_o(3) - 3C_1V_o(2) + 2C_1^3, \\
C_4 &= V_o(4) - 4C_1V_o(3) - 3V_o(2)^2 \\
&\quad + 12C_1^2V_o(2) - 6C_1^4, \\
C_5 &= V_o(5) - 5C_1V_o(4) - 16V_o(2)V_o(3) + 48C_1V_o(2)^2 \\
&\quad + 20C_1^2V_o(3) - 72C_1^3V_o(2) + 24C_1^5 \tag{S.46}
\end{aligned}$$

and

$$\begin{aligned}
D_1 &= 1, \\
D_2 &= \mu_1 - C_1, \\
D_3 &= \mu_2 - V_o(2) - 2C_1\mu_1 + 2C_1^2, \\
D_4 &= \mu_3 - V_o(3) + 6C_1V_o(2) - 3\mu_1V_o(2) + 6\mu_1C_1^2 \\
&\quad - 3\mu_2C_1 - 6C_1^3, \\
D_5 &= \mu_4 - V_o(4) - 4\mu_1V_o(3) + 8C_1V_o(3) - 12\mu_2V_o(2) \\
&\quad + 12V_o(2)^2 + 24C_1\mu_1V_o(2) - 48C_1^2V_o(2) \\
&\quad - 4C_1\mu_3 - 24C_1^1\mu_1 + 12C_1^2\mu_2 + 24C_1^4. \tag{S.47}
\end{aligned}$$

Note that  $C_1$  and  $D_1$  are not related to measured quantities but  $D_1$  is trivially equal to one and  $C_1$  is just the expected number of particles within a ball of radius  $R$ , hence it is just a system parameter.

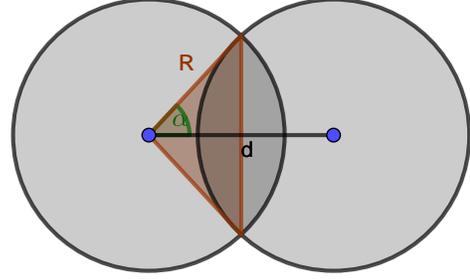


FIG. S1. Overlap area of two disks. It can be calculated by taking twice the difference of areas of the circle segment with angle  $2\alpha$  and of the triangle.

## IV. OVERLAP AREA OF TWO AND THREE DISKS

### A. Two Disks

For two disks of Radius  $R$  and with distance  $d$  between the centers of the disks the overlap area can be calculated as twice the difference of the circle segment and the triangle, see Fig.S1 for a sketch. The area of the circle segment is

$$A_{\text{circlesegment}} = \alpha R^2 \tag{S.48}$$

and the area of the triangle is given by

$$A_{\text{triangle}} = \frac{d}{2} R \sin \alpha, \tag{S.49}$$

where the angle  $\alpha$  is given by

$$\frac{d}{2} = R \cos \alpha. \tag{S.50}$$

Combining these expressions we find for the overlap area

$$A_{\text{overlap}} = 2R^2 \arccos\left(\frac{d}{2R}\right) - \frac{d}{2} \sqrt{4R^2 - d^2}. \tag{S.51}$$

### B. Three Disks

We calculate the overlap area of three disks of radius  $R$  with their centers pairwise separated by distances  $a, b$  and  $c$ . Clearly the overlap area only depends on those three distances. Without loss of generality we assume for simplicity  $a \geq b \geq c$ .

We have to consider three different cases.

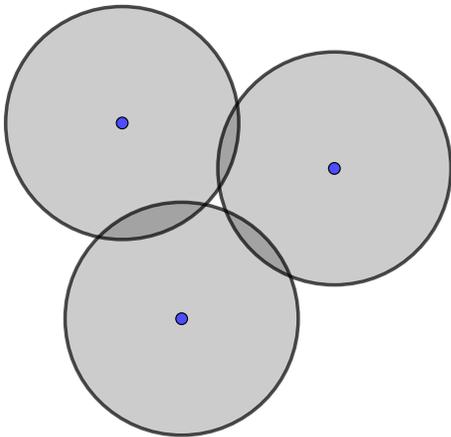


FIG. S2. Case A - no overlap of three disks although there is a pairwise overlap for each pair.

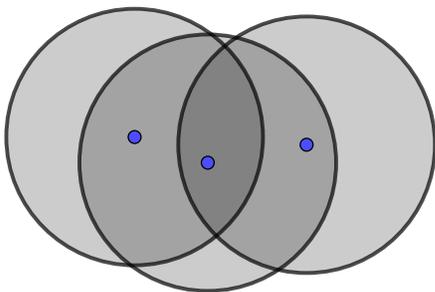


FIG. S3. Case B - the overlap of three disks is identical to the overlap of two of the three disks.

### 1. Case A - no overlap

Clearly, there is no overlap whenever  $a \geq 2R$  since in that case the two circles separated by  $a$  have zero overlap area. However, even if the three disks have a pairwise overlap for all pairs, there might be still zero overlap of all three disks, see Fig.S2. Of course in this case

$$A_{\text{overlap}} = 0. \quad (\text{S.52})$$

### 2. Case B - two disk overlap

It might happen that the three disk overlap is identical to the two disk overlap of the disks separated by  $a$ , see

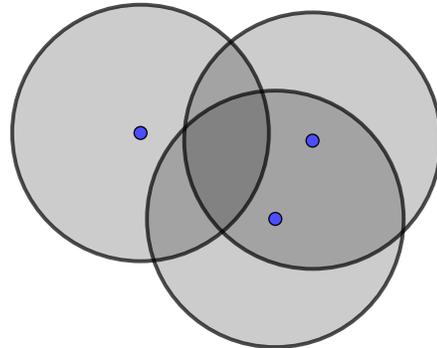


FIG. S4. Case C - overlap of three disks - a generalization of the Reuleaux triangle.

Fig.S3. In that case the overlap area is

$$A_{\text{overlap}} = 2R^2 \arccos\left(\frac{a}{2R}\right) - \frac{a}{2}\sqrt{4R^2 - a^2}, \quad (\text{S.53})$$

which is identical to Eq. (S.51) with  $d$  replaced by  $a$ .

### 3. Case C - Reuleaux triangle like overlap

In this most interesting case the overlap is nonempty and restricted by the boundaries of all three disks, see Fig.S4. In the special case of  $a = b = c = R$  the overlap is a so called Reuleaux triangle which has important applications in arts, architecture and engineering.

Before calculating the overlap area in the third case it is important to find clear criteria, depending only on  $a, b$  and  $c$  that distinguish between the cases A, B and C. Therefore it is useful to study the transitions of  $A \leftrightarrow B$ ,  $A \leftrightarrow C$  and  $B \leftrightarrow C$  by moving one of the disks.

First, we consider the transition from C to A. As soon as the overlap area has fallen to zero we have the situation sketched in Fig.S5. The overlap consists only out of the single point  $S$ . It has the same distance  $R$  to the centers of the three disks. Hence,  $S$  is the center of the circumference of the triangle given by the centers of the disks and the circumference has radius  $R$ . If one of the disks is now moved away from the overlap of the other two there is no overlap at all, we are in case A, and the radius of the circumference increases. If on the other side one of the disks is moved towards the overlap of the other two, the three disks overlap becomes a real area, we are in case C, and the radius of the circumference decreases.

The transition  $A \leftrightarrow B$  occurs when  $a = 2R$ , however, we might just consider the case that  $a < 2R$ , then this transition is not relevant.

The Transition  $B \leftrightarrow C$  is sketched in Fig.S6. Again, the transition occurs when the three circles intersect in

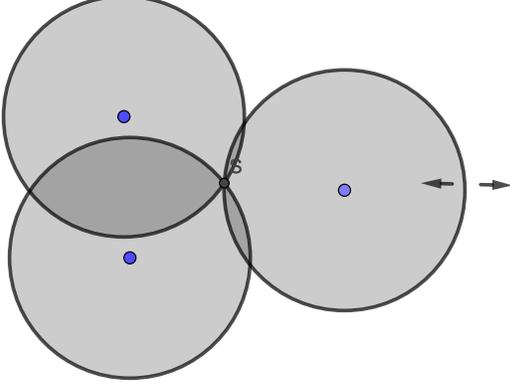


FIG. S5. Transition between cases A and C. The overlap consists only out of the single point  $S$ . It is the center of the circumference of the centers of the disks.

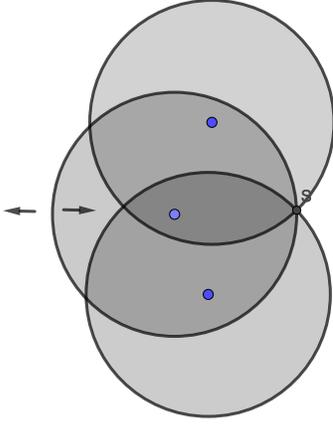


FIG. S6. Transition between cases B and C. The overlap consists only out of the overlap of only two of the disks. The point  $S$  is the center of the circumference of the centers of the disks.

just one point  $S$  which has the same distance  $R$  to the centers of all three disks and thus  $R$  is the circumference diameter. If one of the disks is moved either the overlap becomes just the two circle overlap, we are in case B, and the circumference radius increases or the overlap becomes a Reuleaux-like triangle, we are in case C, and the circumference radius decreases.

Thus we found that case C occurs whenever for the circumference radius it holds

$$r_{\text{circumference}} = \frac{abc}{4\sqrt{s(s-a)(s-b)(s-c)}} < R \quad (\text{S.54})$$

with  $s = (a + b + c)/2$ .

Whenever  $r_{\text{circumference}} \geq R$  we still need to distinguish between cases A and B. As we assumed  $a < 2R$  the midpoint  $M_a$  of the two disk centers separated by  $a$

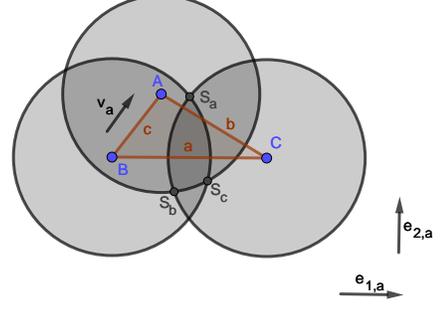


FIG. S7. Sketch of the Reuleaux-like overlap.

is always in the overlap of those two disks. If we draw a circle with diameter  $a$  around  $M_a$  and the center of the third disk is inside this circle it follows that the three-disk overlap is nonempty and thus we are in case B. Thales's theorem tells us that in this case the triangle given by  $a, b, c$  is obtuse. Hence by Pythagoras's theorem we conclude that  $a^2 > b^2 + c^2$  in this case. If on the other hand the center of the third disk lies outside the circle we conclude that  $M_a$  is not in the three-disk overlap and hence we can not be in case B. Thus we must be in case A. Again by Thales's theorem the triangle given by  $a, b, c$  must be acute in this case and therefore by Pythagoras's theorem  $a^2 < b^2 + c^2$ .

In summary, we can distinguish the cases A, B and C by

$$\begin{aligned} \text{Case A} &\leftrightarrow r_{\text{circumference}} \geq R \text{ and } a^2 < b^2 + c^2, \\ \text{Case B} &\leftrightarrow r_{\text{circumference}} \geq R \text{ and } a^2 > b^2 + c^2, \\ \text{Case C} &\leftrightarrow r_{\text{circumference}} < R. \end{aligned} \quad (\text{S.55})$$

It remains to calculate the overlap area in case C. We start to calculate the distances  $l_1, l_2, l_3$  between the corners of the overlap area. We use the cartesian coordinates

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = B, \quad \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = C, \quad \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} = A \quad (\text{S.56})$$

for the centers of the three disks. Hence we find for the distances between the disk centers

$$\begin{aligned} a &= \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \\ b &= \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} \\ c &= \sqrt{(x_1 - x_3)^2 + (y_1 - y_3)^2}. \end{aligned} \quad (\text{S.57})$$

We define the following unit vectors.  $\vec{e}_{1,a}$  points from  $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$  to  $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ ,  $\vec{e}_{1,b}$  points from  $\begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$  to  $\begin{pmatrix} x_3 \\ y_3 \end{pmatrix}$  and  $\vec{e}_{1,c}$  points from  $\begin{pmatrix} x_3 \\ y_3 \end{pmatrix}$  to  $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ . Furthermore we define the vectors  $\vec{v}_i$

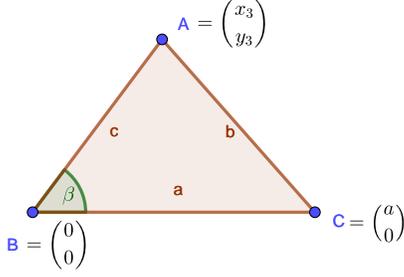


FIG. S8. We choose a coordinate system such that  $B = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ,  $C = \begin{pmatrix} a \\ 0 \end{pmatrix}$  and  $A = \begin{pmatrix} x_3 \\ y_3 \end{pmatrix}$ , with  $x_3, y_3 > 0$ .

via

$$\begin{aligned}\vec{v}_a &= \begin{pmatrix} x_3 - x_1 \\ y_3 - y_1 \end{pmatrix} \\ \vec{v}_b &= \begin{pmatrix} x_1 - x_2 \\ y_1 - y_2 \end{pmatrix} \\ \vec{v}_c &= \begin{pmatrix} x_2 - x_3 \\ y_2 - y_3 \end{pmatrix}.\end{aligned}\quad (\text{S.58})$$

The unit vectors  $\vec{e}_{2,i}$  are defined by being perpendicular to the corresponding unitvector  $\vec{e}_{1,i}$  and the property  $\vec{v}_i \cdot \vec{e}_{2,i} > 0$ , where  $i = a, b, c$ , see Fig. S7 for a sketch of the unit vectors. They can be explicitly calculated by

$$\vec{e}_{2,i} = \frac{\vec{v}_i - (\vec{v}_i \cdot \vec{e}_{1,i})\vec{e}_{1,i}}{|\vec{v}_i - (\vec{v}_i \cdot \vec{e}_{1,i})\vec{e}_{1,i}|}.\quad (\text{S.59})$$

It is a simple geometric consideration to calculate  $S_a$  as

$$S_a = M_a + \sqrt{R^2 - a^2/4}\vec{e}_{2,a},\quad (\text{S.60})$$

where  $M_a$  is the midpoint of the side  $a$ , that is  $M_a = \frac{1}{2}(B + C) = \frac{1}{2}\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}$ . Thus we find the intersection point between the circles  $S_a$  and analogously  $S_b$  as

$$\begin{aligned}S_a &= \frac{1}{2}\begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix} + \sqrt{R^2 - \frac{a^2}{4}}\vec{e}_{2,a}, \\ S_b &= \frac{1}{2}\begin{pmatrix} x_2 + x_3 \\ y_2 + y_3 \end{pmatrix} + \sqrt{R^2 - \frac{b^2}{4}}\vec{e}_{2,b}.\end{aligned}\quad (\text{S.61})$$

We are free to choose a special coordinate system such that  $B = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ ,  $C = \begin{pmatrix} a \\ 0 \end{pmatrix}$  and  $A = \begin{pmatrix} x_3 \\ y_3 \end{pmatrix}$  with  $x_3, y_3 > 0$ , see Fig. S8 for a sketch. Once  $B$  and  $C$  are fixed, the condition  $x_3 > 0$  is one of two choices, whereas the condition  $y_3 > 0$  follows from  $a \geq b \geq c$ . We use the law of cosines to calculate the angle  $\beta = \sphericalangle CBA$  and find

$$\begin{aligned}\cos \beta &= \frac{a^2 + c^2 - b^2}{2ac}, \\ \sin \beta &= \frac{\sqrt{2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4}}{2ac}\end{aligned}\quad (\text{S.62})$$

and hence we found all coordinates of the centers of the disks

$$\begin{aligned}x_1 &= 0 \\ y_1 &= 0 \\ x_2 &= a \\ y_2 &= 0 \\ x_3 &= c \cos \beta = \frac{a^2 + c^2 - b^2}{2a}, \\ y_3 &= c \sin \beta = \frac{\sqrt{2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4}}{2a}.\end{aligned}\quad (\text{S.63})$$

From our choice of the coordinate system we find

$$\begin{aligned}\vec{e}_{1,a} &= \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \\ \vec{e}_{2,a} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix}.\end{aligned}\quad (\text{S.64})$$

As we explicitly have the coordinates of  $C, A$  we can calculate  $\vec{e}_{1,b}$  from its definition and  $\vec{e}_{2,b}$  according to Eq. (S.59) which results in

$$\begin{aligned}\vec{e}_{1,b} &= \frac{1}{2ab} \begin{pmatrix} c^2 - a^2 - b^2 \\ \sqrt{2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4} \end{pmatrix}, \\ \vec{e}_{2,b} &= \frac{1}{2ab} \begin{pmatrix} -\sqrt{2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4} \\ c^2 - a^2 - b^2 \end{pmatrix}.\end{aligned}\quad (\text{S.65})$$

Plugging Eqs. (S.63), (S.64) and (S.65) into Eq. (S.61) we find for two of the intersection points

$$\begin{aligned}S_{a,x} &= \frac{a}{2}, \\ S_{a,y} &= \sqrt{R^2 - \frac{a^2}{4}}, \\ S_{b,x} &= \frac{a}{2} + \frac{a^2 + c^2 - b^2}{4a} \\ &\quad - \sqrt{R^2 - \frac{b^2}{4} \frac{\sqrt{2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4}}{2ab}}, \\ S_{b,y} &= \sqrt{R^2 - \frac{b^4}{4} \frac{c^2 - b^2 - a^2}{2ab}} \\ &\quad + \frac{\sqrt{2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4}}{4a}.\end{aligned}\quad (\text{S.66})$$

Eventually we can calculate the distance between the

points  $S_a$  and  $S_b$  to find

$$\begin{aligned}
l_1 &:= \sqrt{(S_{a,x} - S_{b,x})^2 + (S_{a,y} - S_{b,y})^2} \\
&= \left[ -\frac{\sqrt{2a^2b^2 + 2a^2c^2 + 2b^2c^2 - a^4 - b^4 - c^4}}{2} \right. \\
&\quad \times \left( \frac{\sqrt{R^2 - \frac{a^2}{4}}}{a} + \frac{\sqrt{R^2 - \frac{b^2}{4}}}{b} \right) + \frac{c^2 - b^2 - a^2}{4} \\
&\quad \left. - 2\sqrt{R^2 - \frac{a^2}{4}}\sqrt{R^2 - \frac{b^2}{4}}\frac{c^2 - b^2 - a^2}{2ab} + 2R^2 \right]^{1/2}. \tag{S.67}
\end{aligned}$$

The other two lengths  $l_2$  and  $l_3$  are obtained by permutations of  $a, b$  and  $c$ .

It remains to calculate the overlap area. It consists out of a triangle with side lengths  $l_1, l_2, l_3$  and three circular caps. The triangle area is calculated by Heron's formula, whereas the circular caps can be calculated analogously to the two-particle overlap. Putting everything together we obtain the overlap area

$$\begin{aligned}
\mathcal{A}_{\text{overlap}} &= \arcsin\left(\frac{l_1}{2R}\right) \cdot R^2 - \frac{l_1}{2}\sqrt{R^2 - \frac{l_1^2}{4}} \\
&+ \arcsin\left(\frac{l_2}{2R}\right) R^2 - \frac{l_2}{2}\sqrt{R^2 - \frac{l_2^2}{4}} + \arcsin\left(\frac{l_3}{2R}\right) R^2 \\
&- \frac{l_3}{2}\sqrt{R^2 - \frac{l_3^2}{4}} + \sqrt{l(l-l_1)(l-l_2)(l-l_3)}. \tag{S.68}
\end{aligned}$$

## V. MODELS AND NUMERICAL RESULTS

### A. Vicsek model

In the letter we present numerical results of the standard two dimensional Vicsek model. In this model  $N$  particles move at constant speed  $v$  in individual directions given by angles  $\theta_i$ , that is

$$\begin{aligned}
x_j(t+1) &= x_j(t) + v \cos \theta_j(t), \\
y_j(t+1) &= y_j(t) + v \sin \theta_j(t). \tag{S.69}
\end{aligned}$$

where  $x_j(t), y_j(t)$  define the position of particle  $j$  at time  $t$ . After each unit of time the particles instantaneously change their directions according to the following interaction rule

$$\theta_j(t+1) = \arg \left\{ \sum_{k \in \Omega_j(t+1)} \exp[i\theta_k(t)] \right\} + \eta \xi_i(t), \tag{S.70}$$

where  $\Omega_j$  is the set of indexes of particles that are within distance  $R$  of particle  $j$  such that  $\sqrt{(x_k - x_j)^2 + (y_k - y_j)^2} \leq R$  for all  $k \in \Omega_j$ . We call the particles given by  $\Omega_j$  the neighbors of particle  $j$ . Note

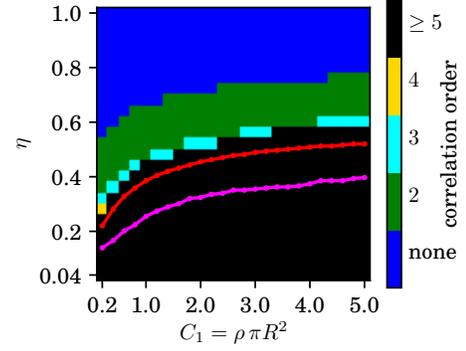


FIG. S9. Minimal required correlation order as in Fig. 2 of the Letter but with a threshold of  $10^{-2}$  for the Kullback-Leibler divergence.

that it is always  $j \in \Omega_j$ . The noise strength is given by  $\eta \in [0, 1]$  and  $\xi_j(t)$  are independent random variables uniformly distributed on the interval  $[-\pi, \pi]$ . It should be noted that the time interval of free motion between the collisions is sometimes considered as a system parameter  $\tau$ . However we only use  $\tau = 1$  as the only relevant parameter is given by the product  $\tau \cdot v$  anyways. Periodic boundary conditions are used for the coordinates  $x_i, y_i \in [-L/2, L/2)$  and  $\phi_i \in [0, 2\pi)$ .

Zero noise  $\eta = 0$  leads to a completely deterministic motion, where particles align consequently and eventually all move in the same direction for almost all initial conditions. On the other hand, all particles move in random directions for  $\eta = 1$ . To distinguish these cases it is useful to introduce the two-dimensional polar order parameter  $\mathbf{p}$  with  $|\mathbf{p}| \in [0, 1]$  defined by

$$p_x = \frac{1}{N} \sum_{j=1}^N \cos \theta_j, \tag{S.71}$$

$$p_y = \frac{1}{N} \sum_{j=1}^N \sin \theta_j. \tag{S.72}$$

Thus  $|\mathbf{p}| = 1$  corresponds to the collective motion of all particles in the same direction and  $|\mathbf{p}| = 0$  means that all particles move independently in random directions.

In Figs. S9 and S10 we show the *correlation map* equivalent to Fig. 2 of the Letter, but obtained using a different threshold for the Kullback-Leibler divergence of  $10^{-2}$  and  $10^{-4}$  in Figs. S9 and S10, respectively. For these other threshold values of the Kullback-Leibler divergence the correlation orders are slightly shifted. However, the over all picture remains unchanged.

In Figs. S11 and S12 we show the three- and four-particle correlation parameters obtained from the same simulations as we used for Fig. 3 of the Letter. They look qualitatively similar to the two-particle correlation parameter  $C_2$ .

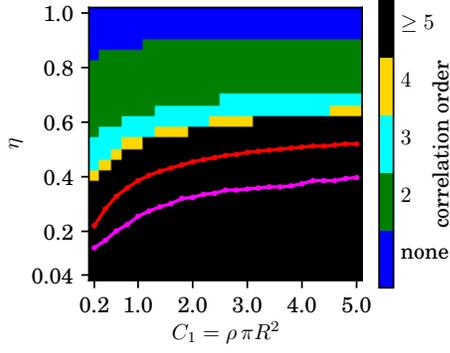


FIG. S10. Minimal required correlation order as in Fig. 2 of the Letter but with a threshold of  $10^{-4}$  for the Kullback-Leibler divergence.

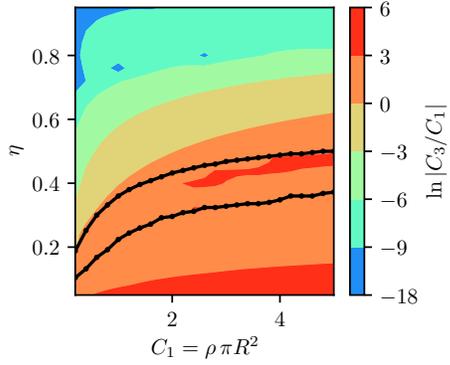


FIG. S11. Ratio of three particle correlation parameter  $C_3$  and mean number of neighbors  $C_1$  for parameters as in Fig. 3 of the Letter. Black lines show the transition lines. For large noise, when  $C_3$  is very small its measured value is sometimes negative (not shown). In principle, the correlation parameters can be negative, however, this is not expected to be the case here due to the aligning interactions. Here, the negative values appear when the measurement uncertainty is larger than the value itself.

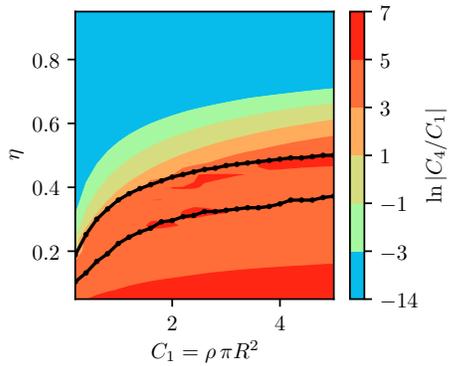


FIG. S12. Ratio of four particle correlation parameter  $C_4$  and mean number of neighbors  $C_1$  for parameters as in Fig. 3 of the Letter. Black lines show the transition lines. For large noise, when  $C_4$  is very small its measured value is sometimes negative (not shown).

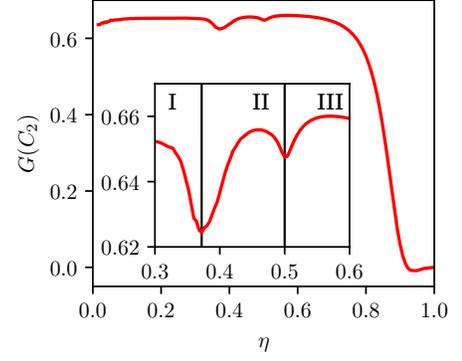


FIG. S13. Binder cumulant of the two-particle correlation parameter  $C_2$  clearly showing two transitions. System parameters:  $N = 22500$ ,  $v\tau/R = 1$ ,  $C_1 = 5$ , thermalization for  $10^5$  and measurement for  $10^6$  time steps for 24 realizations.

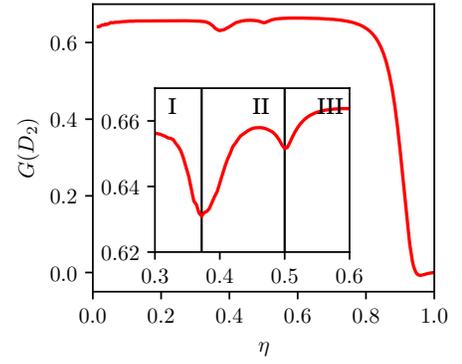


FIG. S14. Binder cumulant of the two-particle correlation parameter  $D_2$  obtained from the same simulations as Fig. S13 also showing the two transitions at the same positions.

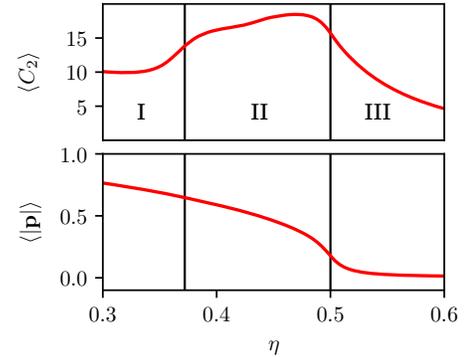


FIG. S15. *Top*: Two-particle correlation parameter  $C_2$ . *Bottom*: Polar order parameter. Data from same simulations as Fig. S13. Vertical lines show transitions obtained as minima of the Binder cumulant in Fig. S13.

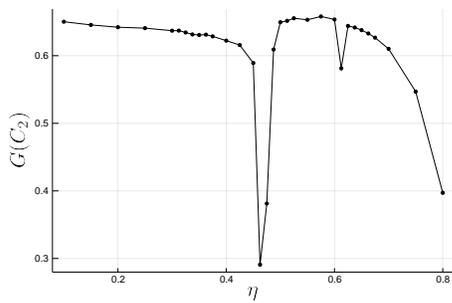


FIG. S16. Binder cumulant of the two-particle correlation parameter  $C_2$  clearly showing two transitions. System parameters:  $N = 22500$ ,  $v\tau/R = 5$ ,  $C_1 = 5$ , thermalization for  $10^5$  and measurement for  $10^6$  time steps for 25 realizations.

In Figs. S13, S14 and S15 we display the Binder cumulants of the two-particle correlation parameters  $C_2$  and  $D_2$ , the mean value of  $C_2$  and the polar order parameter analogously to Fig. 4 of the Letter for another parameter set. Here we see even clearer minima in the Binder cumulant. Furthermore we observe almost no difference in the Binder cumulants of  $C_2$  and  $D_2$ , thus one might equally well use  $D_2$  instead of  $C_2$ . In Fig. S16 we display the Binder cumulant for another parameter set (large speed). There, we also find clearly two transitions.

We obtain the critical point  $\eta_c(L)$  for the polar order transition as the right local minimum of the Binder cumulant in Fig. 4 (c) of the Letter for different finite system sizes  $L$ . To extrapolate to infinitely large systems we use the finite size scaling ansatz

$$\eta_c(L) = \eta_c^\infty - A \cdot L^\alpha, \quad (\text{S.73})$$

If the wave fronts occurring in the polar ordered phase would have the same shape for all system sizes and noise strengths, we would require that the exponent  $\alpha$  equals to minus one. However, the shape of the wave fronts depends on the noise strength. Therefore, we need to fit all three parameters and obtain the values  $\eta_c^\infty = 0.285$ ,  $A = 1.79$  and  $\alpha = -0.711$ . The ansatz (S.73) is shown together with the measured values of  $\eta_c(L)$  in Fig. S17. The results seem to agree very well, however, we should keep in mind that we fitted three parameters with five data points. In order to give a quantitative uncertainty estimate of the fitted parameters, we need data for more and larger system sizes. The results deliver therefore only a rough estimate of the critical noise strength for the infinite system size. However, the value we obtain via finite size scaling is already quite close to the value that was obtained by using a different method (values given in the Letter).

## B. Continuous time Vicsek-like model

Here, we consider a model similar to the Vicsek model but with continuous time. As in the standard Vicsek

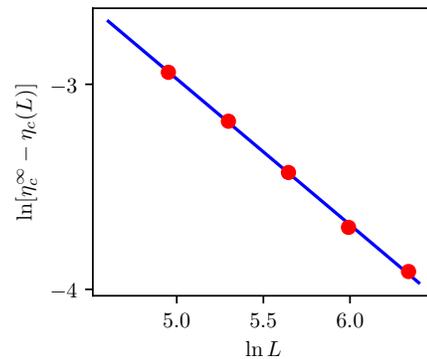


FIG. S17. Finite size scaling for the critical noise strength of the polar order transition of the simulations presented in Fig. 4 of the Letter. The blue line represents a fit of the finite size scaling ansatz (S.73). The red circles show the measured values for different system sizes.

model particles move in two dimensions with constant speed and align their direction of motion with nearby particles. The system is defined by the following Langevin equations

$$\begin{aligned} \dot{x}_i &= v \cos(\theta_i) \\ \dot{y}_i &= v \sin(\theta_i) \\ \dot{\theta}_i &= w(|\Omega(i)|) \sum_{j \in \Omega(i)} \sin(\theta_j - \theta_i) + \sigma \xi_i, \quad i = 1, \dots, N, \end{aligned} \quad (\text{S.74})$$

Here, the neighborhood sets  $\Omega_i$  are defined as in the standard Vicsek model above. The  $\xi_i(t)$  are independent Gaussian white noise terms satisfying

$$\langle \xi_i(t) \xi_j(s) \rangle = \delta_{ij} \delta(t - s). \quad (\text{S.75})$$

The interaction weight  $w$  is a function of the number of neighbors including the particle itself. In this paper we consider the case  $w(n) = 1/n$ .

In Figs. S18 and S19 we show the two-particle and three-particle correlation parameters  $C_2$  and  $C_3$  for the continuous time Vicsek-like model, respectively. The results are qualitatively equivalent to the ones in Fig. 3 of the Letter and Fig. S11.

Note that the computation time of the correlation analysis is of the order of the particle number times the average number of neighbors. This is the same complexity as for the Molecular Dynamics simulation itself. Hence, measuring the correlation parameters does not significantly increase the computation time.

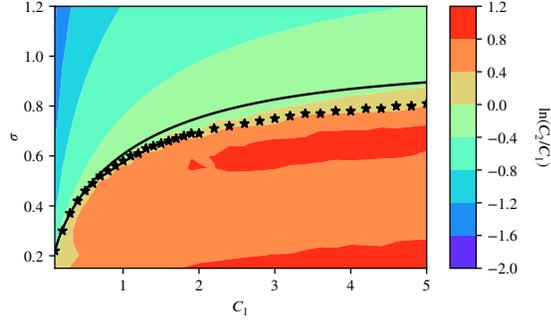


FIG. S18. Logarithm of the two-particle correlation parameter  $C_2$  divided by the average number of neighbors  $C_1$  in dependence on noise strength  $\sigma$  and particle density for the model defined in Eq. (S.74). The black solid line shows the mean field transition towards polar order, whereas the black stars represent the actually measured transition points. Simulation parameters: particle number  $N = 10^5$ , velocity  $v = 1$ , interaction radius  $R = 1$ , time step  $\Delta t = 0.03$ , the thermalization time as well as the measurement time was  $5 \times 10^4$  time units. The results have been averaged over 16 realizations.

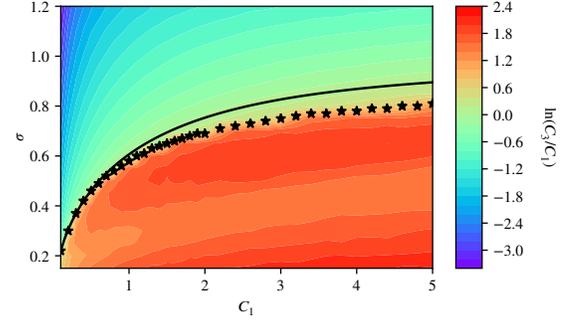


FIG. S19. Logarithm of the three-particle correlation parameter  $C_3$  divided by the average number of neighbors  $C_1$  in dependence on noise strength  $\sigma$  and particle density for the model defined in Eq. (S.74). The black solid line shows the mean field transition towards polar order, whereas the black stars represent the actually measured transition points. Parameters as in Fig. S18.

---

[S1] M. Kerscher, in *Statistical Physics and Spatial Statistics*, edited by K. R. Mecke and D. Stoyan (Springer, Berlin, Heidelberg, 2000) pp. 36–71.