

Cluster Formation in Stochastic Disk Systems

G. Sutmann^{1,2,a)}, H. Ganesan² and C. Begau²

¹*Jülich Supercomputing Centre (JSC), Institute for Advanced Simulation (IAS),
Forschungszentrum Jülich (JSC), D-52425 Jülich, Germany*

²*Interdisciplinary Centre for Advanced Materials Simulation, ICAMS, Ruhr-University Bochum, Germany*

^{a)}Corresponding author: g.sutmann@fz-juelich.de

Abstract. The problem of randomly distributed disks is considered in the dilute regime in a two-dimensional domain. Disks are allowed to overlap and to form clusters which may be isolated or percolating. Depending on the number and size of the disks, distribution functions are obtained for different size and bond configurations of clusters. A statistical geometrical approach is taken to derive analytical probabilities for cluster formation in systems, where a maximum of four overlapping disks is considered. Monte Carlo computations are carried out to verify our theoretical approach which is shown to be in close agreement with numerical simulations.

INTRODUCTION

The present work was motivated by the analysis of Monte Carlo simulations of many-particle systems. Considering N particles which interact via a potential $U(\mathbf{r})$, leads in general to an $O(N^2)$ problem. If the potential function drops down sufficiently fast, interactions can be truncated after a characteristic region Ω_{R_c} with radius R_c . Since number of particles within Ω_{R_c} is bound, the problem can be reduced to $O(N)$ [1]. In Monte Carlo simulations of many-particle systems, trial moves are often conducted by single particle displacements [2]. Acceptance of trial moves is decided upon an acceptance criterion, e.g. the Metropolis algorithm [2]. Including relaxation of all particles within $\Omega_{R_c}(\{i\})$ implies configurational changes. Considering a subset of tagged particles $\{i\} = \{i_1, \dots, i_n\}$ with their proper regions $\Omega_{R_c}(\{i\})$, the evaluation of energies of tagged particles and independent relaxation could be performed independently if $\Omega_{R_c}(i_{k_1}) \cap \Omega_{R_c}(i_{k_2}) = \emptyset$, $\forall k_1, k_2 \in \{1, n\}$ and therefore would allow for parallel execution. However, usually the tagged particles are chosen stochastically with a probability of $p_i = 1/N$ to avoid a bias in the computation and therefore there is a non-vanishing probability for $\Omega_{R_c}(i_{k_1}) \cap \Omega_{R_c}(i_{k_2}) \neq \emptyset$, $\exists k_1, k_2 \in \{1, n\}$, which creates a causal dependence between the energy computations. A possible solution to this problem is to discard completely the region which was chosen latest and which causes an overlap or to perform a speculative execution and to consider both trial moves as valid Monte Carlo attempts, if at least one of them is rejected in the energy criterion. Only if both trials would be accepted, the overlapping region should have to be discarded completely.

The present work provides an analysis of such scenarios, where regions $\Omega_{R_c}(\{i\})$ are distributed within a finite domain. Probabilities are derived for the occurrence of clusters of different size and bonding configurations. Since the overlap between regions is dependent on the central tagged particle and the size of the interaction range with radius R_c , this problem can be reduced to the one of distribution of disks with radius R_c in a finite domain. Analysis is provided for diluted systems, i.e. ones for which the cluster size can be considered to be bound for all practical purposes. A complete analysis for an N -particle system would be practically impossible (e.g. for $N = 10$ the largest cluster would allow for more than 10^7 non-isomorphic graphs [3]). We derive probabilities for the appearance of clusters by geometric considerations, which are based on average disk configurations.

In the present article we restrict ourselves to periodic boundary conditions. Probabilities for 2-particle distances in other boundary conditions are known as line picking problems [4], which are, however not easy to extend to larger cluster configurations, i.e. $N > 2$. Overlap in disk systems has been considered in the context of percolation in random systems and was first considered by Penrose [5] for Poisson processes and which considers probabilities for forming bonds between nearby particles. Later on this concept was further exploited to study percolation in penetrable disks,

spheres [6, 7] and ellipsoids [8]. Clusters are thereby computed as ensemble averages of random particle distributions.

Here we follow a somewhat different approach by considering probabilities of formation of penetrable disk clusters in terms of graph realizations. Average geometries are considered and overlap probabilities of a particle with a $(k-1)$ -cluster are obtained by considering effective cross-sections of the cluster to form a k -cluster. Here we consider the case up to $k = 4$ and probabilities of formation of all connected and disconnected graphs up to order 4 can be obtained as analytical expressions. Schematically we describe the procedure and present results, which are compared with Monte Carlo computations of cluster statistics, which are in very good agreement with theoretical predictions in the low density regime. This provides single cluster probabilities which can be used in a next step for the statistical analysis of larger systems, where combinatorial rules will be applied and which will be presented in another article [9].

OVERLAP PROBABILITIES

We consider disks of radius R_c , randomly placed inside a region of size $\Omega_s = [0, 1]^2$ under periodic boundary conditions. The computation of the overlap probability between disks can be mapped to the random line picking problem [4], where the distance between two randomly placed points in space is considered. If the distance d between points is $d < 2R_c$, disks overlap and form a 2-cluster (a cluster, which is formed by n particles is henceforth called an n -cluster). If a first particle is placed randomly inside Ω_s and a second particle is placed at a random position, where variates $x, y \in [0, 1]$ are drawn from a uniform distribution, it is obvious that the probability, i.e. the geometric cross-section, for hitting an area element $\{r, r + dr\}$ and $\{\varphi, \varphi + d\varphi\}$ at a distance r from a point is $p(\varphi, r) dr d\varphi = r dr d\varphi$ and therefore the total probability to form a 2-cluster is $P_2 = \int_0^{2\pi} d\varphi \int_0^{2R_c} dr r = 4\pi R_c^2$ which is the area of a disk with radius $2R_c$.

To obtain the probability of larger clusters, we have to distinguish different cluster configurations for a given number of particles. In general, this can be obtained by graph theoretical considerations, if we consider every particle as a vertex and an overlap between particles as a bond or an edge in the graph. Our goal here is to derive probabilities for clusters up to size of $N = 4$. From graph theory it is known that for $N = 2$ there is only one type of cluster if we consider only the connected graphs. For $N = 3$ there are 2 types of graphs, one with two and one with three edges, where the one with two edges has three different permutation states. For $N = 4$ there are 6 connected graphs, 2 with three edges, 2 with four edges, 1 with five edges and 1 with six edges. Before considering these cases in more detail, we compute the average configuration of a 2-cluster, i.e. the average distance between particles

$$\langle d \rangle = \frac{1}{4\pi R_c^2} \int_0^{2\pi} d\varphi \int_0^{2R_c} dr r^2 = \frac{4}{3} R_c \quad (1)$$

From elementary geometry it is found that the overlap area $A_o(R, d)$ between two disks of radius R in a distance d is given by $A_o(R, d) = 2R^2 \cos^{-1}(d/2R) - d/2 \sqrt{4R^2 - d^2}$ so that the average overlap area over an interval $r \in [0, 2\pi]$ for disks with cross-section $2R_c$ is obtained as

$$\langle A_o(2R_c) \rangle_d = \frac{1}{4\pi R_c^2} \int_0^{2\pi} d\varphi \int_0^{2R_c} dr \left[8r R_c^2 \cos^{-1}\left(\frac{r}{4R_c}\right) - \frac{r^2}{2} \sqrt{16R_c^2 - r^2} \right] = (4\pi - 3\sqrt{3}) R_c^2 \quad (2)$$

To compute the probabilities of cluster formation of higher order, geometric criteria can be considered. In Figure 1 the possible graphs for a k -cluster are illustrated together with the eligible area $\bar{\Omega}_k$ within a $(k-1)$ -cluster, which has to be hit by a k th particle to form the proper k -cluster. The probability is then computed as the $(k-1)$ -cluster probability multiplied by $\bar{\Omega}_k$.

Probabilities for 3-clusters

Formation of (\mathcal{A}): The eligible area for a third particle is given as $\bar{\Omega}_3 = \Omega_2(2R_c) - \Omega_1(2R_c) \cap \Omega_2(2R_c)$, which is $\bar{\Omega}_{3,1} = 4\pi R_c^2 - \langle A_o(2R_c) \rangle_d$, and therefore $P_{3,1} = P_2 \times \bar{\Omega}_{3,1} = 12\sqrt{3}\pi R_c^4$.

Formation of (\mathcal{B}): the eligible area $\bar{\Omega}_{3,2}$ is calculated as the average overlap between two disks of radius $2R_c$ and therefore $P_{3,2} = P_2 \times \bar{\Omega}_{3,2} = 4\pi(4\pi - 3\sqrt{3}) R_c^4$.

Probabilities for 4-clusters

The probabilities for 4-clusters, represented by their proper graphs (cmp. Figure 1) are given in summary. We have derived closed form analytical expressions for the eligible areas, which are partly too complex for the present article and will be presented in full detail elsewhere [9].

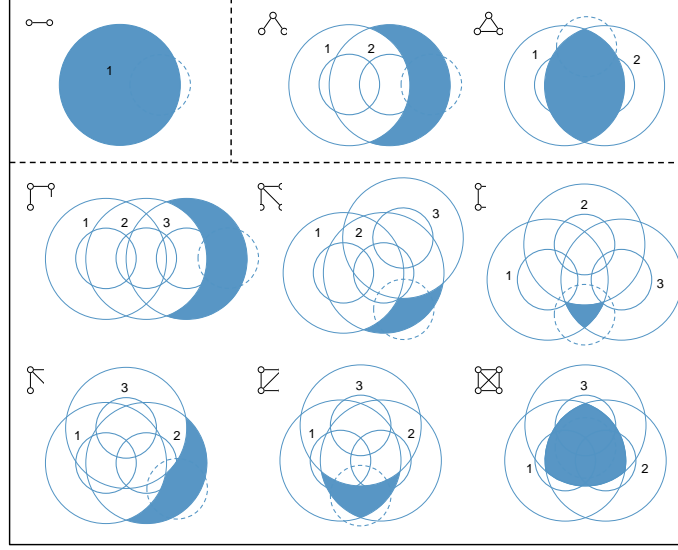


FIGURE 1. Cluster configurations of size $N - 1$. Small circles represent disks of radius R_c , which overlap in their average configuration. Large circles of size $2R_c$ represent area of potential overlap if another particle is added. The shaded area indicates eligible area, where an additional particle has to be positioned to form a cluster represented by the indicated graph. For clarity a possible k th particle is shown as dotted circle.

- Formation of $\text{---}\circ\text{---}\circ$: $P_{4,1} = P_{3,1} \times \bar{\Omega}_{4,1}$, with $\bar{\Omega}_{4,1} = \Omega_3 - \Omega_2 \cap \Omega_3$
- Formation of $\text{---}\circ\text{---}\circ\text{---}\circ$: $P_{4,2} = P'_{3,1} \times \bar{\Omega}_{4,2}$, with $\bar{\Omega}_{4,2} = \Omega_3 - \Omega_1 \cap \Omega_3 - \Omega_2 \cap \Omega_3 + \Omega_1 \cap \Omega_2 \cap \Omega_3$
- Formation of $\text{---}\circ\text{---}\circ\text{---}\circ$: $P_{4,3} = P''_{3,1} \times \bar{\Omega}_{4,3}$, with $\bar{\Omega}_{4,3} = \Omega_1 \cap \Omega_3 - \Omega_1 \cap \Omega_2 \cap \Omega_3$
- Formation of $\text{---}\circ\text{---}\circ\text{---}\circ$: $P_{4,4} = P_{3,2} \times \bar{\Omega}_{4,4}$, with $\bar{\Omega}_{4,4} = \Omega_3 - \Omega_1 \cap \Omega_3 - \Omega_2 \cap \Omega_3 + \Omega_1 \cap \Omega_2 \cap \Omega_3$
- Formation of $\text{---}\circ\text{---}\circ\text{---}\circ$: $P_{4,5} = P_{3,2} \times \bar{\Omega}_{4,5}$, with $\bar{\Omega}_{4,5} = \Omega_1 \cap \Omega_3 - \Omega_1 \cap \Omega_2 \cap \Omega_3$
- Formation of $\text{---}\circ\text{---}\circ\text{---}\circ$: $P_{4,6} = P_{3,2} \times \bar{\Omega}_{4,6}$, with $\bar{\Omega}_{4,6} = \Omega_1 \cap \Omega_2 \cap \Omega_3$

It is obvious that the eligible area of $P_{4,2}$ and $P_{4,4}$ is computed in a similar way. However, the underlying average geometry of particles 1, 2, 3 is different in both cases and so the individual $\bar{\Omega}$'s are different. The same is true for $P_{4,3}$ and $P_{4,5}$, where the probabilities $P'_{3,1}$, $P''_{3,1}$ restricts the formation of the underlying 3-cluster to geometries which allow for a formation of the 4-clusters ($\text{---}\circ\text{---}\circ\text{---}\circ$, $\text{---}\circ\text{---}\circ\text{---}\circ$).


















RESULTS

We have computed the individual probabilities of cluster formation according to the prescription given in the previous Section for a set of particle sizes. Average configurations of clusters are approximated in the following way: the distance between 2 disks is taken as the average overlap distance, i.e. $\langle d \rangle = 4/3R_c$. For the higher order clusters, structures are constructed by symmetry arguments. E.g. 3-clusters, represented by the two connected 3-graphs are built by a linear configuration for $\text{---}\circ\text{---}\circ$ and an equilateral triangular configuration for $\text{---}\circ\text{---}\circ$ clusters, where distances between disks are taken as average distance $\langle d \rangle$ (cmp. Figure 1). For the 4-clusters a similar procedure is followed. Only in the case of $\text{---}\circ\text{---}\circ$, the average distance $\langle d_4 \rangle = (8 - \sqrt{8})/3R_c$ between two disks is computed from the interval $\{\sqrt{2}R_c, 2R_c\}$ which is in agreement with the connected 4-graph. Probabilities p of configurations, represented by not connected graphs (e.g. $\text{---}\circ\text{---}\circ$), are computed as probability to form a cluster times the probability to not form a cluster with other particles.

Explicit theoretical results are listed in Table 1. As reference, we have conducted Monte Carlo simulations, where 10^9 trial configurations were analyzed for systems of $N \in \{2, 4\}$ in a system with area $\Omega_s = [0, 1]^2$. All distinct clusters, represented by different graphs were identified and their probability of formation was computed by geometrical integration. It is found that most theoretical predictions are in very good agreement with Monte Carlo

results. This applies especially to the cases of small disks. For the case of $R_c = 0.2$ finite size effects are expected where the extent of the clusters gets larger than the box length, which is a signature of percolation.

TABLE 1. Overlap probabilities between circles in a $[0, 1]^2$ box. Compared are results from Monte Carlo evaluations (p_{MC}) and theoretical predictions (p_{th}), based on effective overlap areas between circles. Shown are results for systems with $N = 2, 3, 4$ and different realizations of bonding configurations, represented by their proper graphs \mathcal{G} .

N	\mathcal{G}	p_{th}				p_{MC}			
		0.05	0.10	0.15	0.2	0.05	0.10	0.15	0.2
2		0.9686	0.8743	0.7172	0.4973	0.9686	0.8743	0.7173	0.4973
		0.0314	0.1257	0.2828	0.5027	0.0314	0.1257	0.2827	0.5027
3		0.9087	0.6684	0.3690	0.1230	0.9081	0.6611	0.3447	0.0946
		0.0300	0.1032	0.1689	0.1427	0.0300	0.1033	0.1698	0.1528
		4.08×10^{-4}	6.53×10^{-3}	3.31×10^{-2}	1.04×10^{-1}	4.08×10^{-4}	6.53×10^{-3}	3.31×10^{-2}	9.72×10^{-2}
		5.79×10^{-4}	9.26×10^{-3}	4.69×10^{-2}	1.48×10^{-1}	5.79×10^{-4}	9.26×10^{-3}	4.69×10^{-2}	1.55×10^{-1}
4		0.8257	0.4468	0.1362	0.0151	0.8237	0.4270	0.1007	0.0053
		2.77×10^{-2}	7.41×10^{-2}	7.23×10^{-2}	2.02×10^{-2}	2.77×10^{-2}	7.30×10^{-2}	6.56×10^{-2}	1.39×10^{-2}
		9.43×10^{-4}	1.30×10^{-2}	4.77×10^{-2}	7.18×10^{-2}	9.28×10^{-4}	1.20×10^{-2}	3.67×10^{-2}	3.45×10^{-2}
		3.84×10^{-4}	4.97×10^{-3}	1.53×10^{-2}	4.47×10^{-3}	3.84×10^{-4}	4.96×10^{-3}	1.53×10^{-2}	1.39×10^{-2}
		5.48×10^{-4}	7.29×10^{-3}	2.44×10^{-2}	2.17×10^{-2}	5.50×10^{-4}	7.40×10^{-3}	2.56×10^{-2}	3.42×10^{-2}
		5.34×10^{-6}	3.42×10^{-4}	3.89×10^{-3}	2.19×10^{-2}	5.75×10^{-6}	3.69×10^{-4}	4.14×10^{-3}	1.38×10^{-2}
		1.81×10^{-6}	1.16×10^{-4}	1.32×10^{-3}	7.40×10^{-3}	1.81×10^{-6}	1.17×10^{-4}	1.33×10^{-3}	5.79×10^{-3}
		1.91×10^{-7}	1.22×10^{-5}	1.40×10^{-4}	7.84×10^{-4}	1.71×10^{-7}	1.08×10^{-5}	1.87×10^{-4}	6.83×10^{-3}
		4.90×10^{-6}	3.14×10^{-4}	3.57×10^{-3}	2.01×10^{-2}	4.11×10^{-6}	2.63×10^{-4}	2.99×10^{-3}	1.48×10^{-2}
		2.67×10^{-6}	1.71×10^{-4}	1.95×10^{-3}	1.09×10^{-3}	2.78×10^{-6}	1.78×10^{-4}	2.03×10^{-3}	1.34×10^{-2}
		7.94×10^{-6}	5.08×10^{-4}	5.79×10^{-3}	3.25×10^{-2}	8.44×10^{-6}	5.44×10^{-4}	6.20×10^{-3}	3.65×10^{-2}

DISCUSSION

We have presented explicit results for the formation of all possible configurations of small clusters in 2-dimensional disk systems. The work was motivated by the analysis of parallel Monte Carlo simulations, which suffer in efficiency when areas of influence, administered by different threads show overlap. Taking the fully disconnected graphs as cases where no performance degradation is to be expected we see that for small regions of influence with diameters $2R_c$ of about 10% of the system dimension, the efficiency will be larger than 80%, which represents a conservative lower limit for the parallel efficiency. This will of course change when more than 4 threads are used, i.e. more than 4 regions of influence are concurrently computed. However, a full analysis will also consider cases with overlap between regions which will be presented in a subsequent publication [9].

REFERENCES

- [1] G. Sutmann and V. Stegailov, J. Mol. Liq. **125**, 197–203 (2006).
- [2] D. Frenkel and B. Smit, *Understanding molecular simulation. From algorithms to applications*. (Academic Press, San Diego, 2002).
- [3] J. Novak, “Three lectures on free probability,” in *Random Matrix Theory, Interacting Particle Systems and Integrable Systems*, MSRI Publications 65, edited by P. Deift and P. Forrester (Cambridge University Press, Cambridge, 2014), pp. 309–383.
- [4] E. Weisstein, Mathworld, <http://mathworld.wolfram.com/topics/RandomPointPicking.html>.
- [5] M. Penrose, Adv. Appl. Prob. **23**, 536–556 (1991).
- [6] J. Quintanilla and S. Torquato, Phys. Rev. E **54**, 5331–5339 (1996).
- [7] J. Quintanilla and S. Torquato, J. Phys. A: Math. Gen. **33**, L399–L407 (2000).
- [8] Y.-B. Yi and A. Sastry, Phys. Rev. E **66**, p. 066130 (2002).
- [9] G. Sutmann, H. Ganesan and C. Begau (in preparation).
- [10] K. Esselink, L. D. J. C. Loyens, and B. Smit, Phys. Rev. E **51**, 1560–1568 Feb (1995).