

Principal component analysis of the excluded area of two-dimensional hard particles

Cite as: J. Chem. Phys. **150**, 184906 (2019); <https://doi.org/10.1063/1.5092865>

Submitted: 15 February 2019 . Accepted: 22 April 2019 . Published Online: 14 May 2019

Thomas Geigenfeind, and Daniel de las Heras 



View Online



Export Citation



CrossMark

The Journal
of Chemical Physics

2018 EDITORS' CHOICE

READ NOW!



Principal component analysis of the excluded area of two-dimensional hard particles

Cite as: J. Chem. Phys. 150, 184906 (2019); doi: 10.1063/1.5092865

Submitted: 15 February 2019 • Accepted: 22 April 2019 •

Published Online: 14 May 2019



View Online



Export Citation



CrossMark

Thomas Geigenfeind and Daniel de las Heras^{a)} 

AFFILIATIONS

Theoretische Physik II, Physikalisches Institut, Universität Bayreuth, D-95440 Bayreuth, Germany

^{a)}delasheras.daniel@gmail.com. URL: www.danieldelasheras.com

ABSTRACT

The excluded area between a pair of two-dimensional hard particles with given relative orientation is the region in which one particle cannot be located due to the presence of the other particle. The magnitude of the excluded area as a function of the relative particle orientation plays a major role in the determination of the bulk phase behavior of hard particles. We use principal component analysis (PCA) to identify the different types of excluded areas corresponding to randomly generated two-dimensional hard particles modeled as non-self-intersecting polygons and star lines (line segments radiating from a common origin). Only three principal components are required to have an excellent representation of the value of the excluded area as a function of the relative particle orientation for sufficiently anisotropic particles. Independent of the particle shape, the minimum value of the excluded area is always achieved when the particles are antiparallel to each other. The property that affects the value of the excluded area most strongly is the elongation of the particle shape. PCA identifies four limiting cases of excluded areas with one to four global minima at equispaced relative orientations. We study selected particle shapes using Monte Carlo simulations.

Published under license by AIP Publishing. <https://doi.org/10.1063/1.5092865>

I. INTRODUCTION

Hard body models, for which the interaction potential is infinite if two particles overlap and zero otherwise, are excellent candidates to model colloidal particles, which are dominated by excluded volume. Hard body models are also relevant to understand how the microscopic properties at the particle level determine the macroscopic properties of the system such as its bulk phase behavior. Since the then unexpected classical result of fluid-solid phase transition in a system of hard spheres,¹ the bulk phase behavior of several three-dimensional hard body models has been analyzed by both computer simulations and theoretical approaches such as density functional theory. Anisotropic hard particles form a surprisingly rich variety of mesophases such as uniaxial, biaxial, and cubatic nematics, as well as cholesteric, smectic, and columnar phases. We refer the reader to Ref. 2 for a recent review on phase equilibria of hard body models.

The phase behavior of several two-dimensional hard models has been also reported in the literature. Examples are hard disks,^{3–5} needles,⁶ rectangles,^{7–9} discorectangles,^{7,10,11} triangles,^{12,13} squares,¹⁴ rounded squares,¹⁵ pentagons,¹⁶ hexagons,^{17,18} ellipses,¹⁹ zigzags,²⁰ hockey sticks,²¹ bananalike,²² and allophiles.²³

Phase transitions in hard bodies are driven by entropy. At sufficiently low density, the entropy of the ideal gas dominates, and the system remains isotropic with neither positional nor orientational order. As the density increases, excluded volume effects become more important and the gain in configurational entropy can drive a transition to a phase with orientational and/or positional order.

The study of excluded volume effects in hard bodies starts with the properties of the excluded volume between two particles, which plays a role similar to the pair interaction potential in soft systems (i.e., with continuous interaction potentials). Here, we restrict ourselves to two-dimensional particles. As a result of the potential being infinite if two particles overlap, there exists around each particle an exclusion region in which no other particle can be located. The phase behavior of the system is determined by the ideal gas entropy and the properties of this complicated many-body exclusion region, which depends on the positions and orientations of all particles in the system. Within a mean fieldlike approach, the properties of the many-body exclusion region are fully characterized by the excluded area, which is the area inaccessible to one particle due to the presence of another particle. The excluded area alone does not determine

the complete phase behavior of the system, but it plays a vital role in the determination of the type of stable phases. The value of the excluded area, A_{exc} , depends on the particle shape. For a given particle shape, A_{exc} is a function of the relative orientation between the two particles, ϕ . For example, the value of the excluded area between two rectangles is minimum if their relative angle ϕ is either 0 (parallel) or π (antiparallel). For squares, the value of the excluded area is minimum if $\phi = 0, \pi/2, \pi,$ and $3\pi/2$. Although different particle shapes might generate different excluded areas, it is evident that not any function $A_{\text{exc}}(\phi)$ corresponds to a valid particle shape (e.g., the value of the excluded area cannot be negative). We investigate here the admissible shapes of the function $A_{\text{exc}}(\phi)$. To this end, we apply principal component analysis^{24,25} (PCA) to a set of excluded areas corresponding to randomly generated two-dimensional hard bodies. PCA is an unsupervised learning algorithm intended to simplify the complexity of a high-dimensional data set. Mathematically, PCA is an orthogonal transformation of the data to a new basis in which the basis vectors are sequentially chosen such that the variance of the projection of the original data onto the new basis vectors is as large as possible. As a result of our PCA, we identify four special or limiting types of excluded area with one to four global minima as a function of the relative orientation. Finally, we perform Monte Carlo (MC) simulations for selected particle shapes.

II. METHODS

We aim to find the relevant types of excluded areas $A_{\text{exc}}(\phi)$ in two dimensions. To this end, we first generate random hard particles, and then apply principal component analysis to the excluded area between two identical particles.

A. Particle generation

We generate two-dimensional hard particles following either of two procedures. In the first one, each particle is modeled as a simple (non-self-intersecting) polygon, as sketched in panels (a) and (b) of Fig. 1. A polygon is defined by a closed and ordered set of m random vertices connected via straight edges. For each polygon, the

coordinates of each vertex $\mathbf{x}_i, i = 1, \dots, m$ are obtained by sampling from uniform distributions a random radius $r_i \in (0, 1]$ and a random angle $\phi_i \in [0, 2\pi]$. We allow both convex and concave shapes, but restrict the set to simple, non-self-intersecting, polygons, i.e., the line segments connecting the vertices of a polygon are not allowed to intersect (apart from the end points of two neighboring edges that are joint at one vertex). Furthermore, no two vertices have the exact same coordinates. If the random creation of vertices of a polygon leads to self-intersections, we perform Lin 2-opt moves until all intersections are removed. The algorithm was initially proposed to solve the salesman problem,^{26,27} and it works as follows: Assume the line segment connecting vertices \mathbf{x}_{i-1} and \mathbf{x}_i and the line segment connecting vertices \mathbf{x}_{j-1} and \mathbf{x}_j intersect ($j - i > 1$ and periodic vertices). Then, we reverse the order of all vertices between i and j , i.e., the sequence $\{i - 1, i, \dots, j - 1, j\}$ becomes $\{i - 1, j - 1, \dots, i, j\}$. We perform as many Lin 2-opt moves as required to remove all self-intersections.

To increase the spectrum of particle shapes, we also generate particles modeled as star lines; see Fig. 1(c). A star line has a center \mathbf{x}_0 connected via line segments to m vertices located at $\mathbf{x}_i, i = 1, \dots, m$. The simple polygons contain the star lines as a limiting case of polygons with zero area. However, it is very unlikely to generate a star line with the method we use to generate polygons.

As a final step, each particle (polygons and star lines) is rescaled to the maximum possible size that fits in a square bounding box of length $h = 1$, which defines our unit of length.

B. Computation of the excluded area

In what follows and for simplicity, we refer to $A_{\text{exc}}(\phi)$ as simply the excluded area. To calculate the excluded area between two particles (let them be polygons or star lines) at a given relative orientation ϕ , we use the fact that whenever two identical particles overlap then at least two edges intersect. We first fix the relative orientation between two particles ϕ , and the position of particle 1 (at the origin).

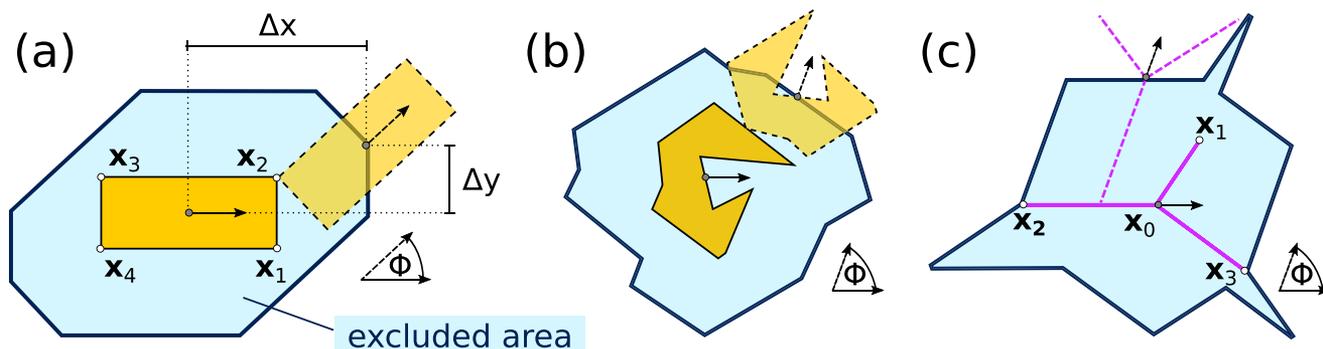


FIG. 1. Excluded area. (a) A rectangle (solid orange) with vertices (empty circles) at $\mathbf{x}_i, i = 1, 2, 3, 4$ and a second identical rectangle (dashed orange) shifted by $(\Delta x, \Delta y)$ and rotated by ϕ with respect to the first particle. The blue area is a graphical representation of the excluded area for the center of the second rectangle at this relative orientation. The black arrow indicates the orientation of the particles (with respect to an arbitrary axis). (b) Graphical representation of the excluded area (at a given relative orientation) of a non-self-intersecting polygon with 10 randomly generated vertices. (c) Excluded area of a randomly generated star line (magenta) with three vertices \mathbf{x}_i connected to the point \mathbf{x}_0 .

Next, we fix the y -distance between the two particles Δy and calculate the interval(s) in x for which the two particles overlap. We loop over all pairs of edges between particles 1 and 2 checking for which x -distances between the particles Δx the edges overlap. The distance Δx (Δy) is calculated as the separation in \hat{x} (\hat{y}) between two reference points located in the particles such as, e.g., the centers of masses. For a fixed value of Δy , a pair of edges intersects either not at all or in at most one connected interval of Δx . The bounds of this interval represent the cases for which a vertex of one edge lies on top of the other edge. Hence, the interval in Δx for which two edges overlap can be obtained by checking the values of Δx for which each of the four involved vertices lies on top of the other edge. Combining the overlapping intervals of all possible pairs of edges gives the complete overlap in the x -direction for a given value of Δy .

Next, we move particle 2 in the y -direction in discrete steps of size $\Delta/h \approx 5 \times 10^{-4}$ between $\Delta y_{\min}(\phi)$ and $\Delta y_{\max}(\phi)$ which are the minimum and maximum values of Δy for which overlap is possible, respectively. Δy_{\min} (Δy_{\max}) occurs when the vertex with the highest (lowest) y -coordinate of particle 2 and that with the lowest (highest) y -coordinate of particle 1 have the same y -coordinate.

Integrating the overlapping intervals between Δy_{\min} and Δy_{\max} yields the excluded area for the selected orientation. Finally, we repeat the process for $n_\phi = 360$ relative orientations $\phi \in [0, 2\pi]$ between the particles and normalize the excluded area according to

$$\frac{1}{2\pi h^2} \int_0^{2\pi} A_{\text{exc}}(\phi) d\phi = 1. \quad (1)$$

C. Principal component analysis

We apply principal component analysis^{24,25} to the excluded areas $A_{\text{exc}}(\phi)$ generated by the above method. All data are organized in a data matrix \mathbf{X} . Each row contains one sample, i.e., the excluded area as a function of the relative orientation for one randomly generated particle. In each column, we store the values of the excluded areas at a given relative orientation for all samples in the system. First, the data are centered to facilitate the following calculations. That is, from each column of \mathbf{X} , the mean is subtracted. We then apply the PCA algorithm. PCA uses an orthogonal transformation to represent the data in a new orthogonal basis. In this basis, the first basis vector \mathbf{w}_1 (first principal axis) is chosen such that the variance of the projection of the data onto this vector is as large as possible, that is,

$$\mathbf{w}_1 = \arg \max_{\|\mathbf{w}_1\|=1} \{\|\mathbf{X}\mathbf{w}_1\|^2\} = \arg \max_{\|\mathbf{c}_1\|=1} \{\|\mathbf{c}_1\|^2\}. \quad (2)$$

Here, the i th component of the vector \mathbf{c}_1 is the first principal component of the i th sample. The variance of the following basis vectors, \mathbf{w}_j with $j > 1$, is also maximized under the constraint that each vector is orthogonal to the preceding ones. The new basis vectors are called principal axes, and the components of a vector expressed in this basis are the principal components. It is often the case that only the first principal components have high variance and are therefore relevant to describe the data. Hence, PCA allows a meaningful dimension reduction, while retaining as much information as possible.

Mathematically, the principal axes are the (normalized) eigenvectors of the covariance matrix \mathbf{C} of the data, which is defined as

$$\mathbf{C} = \frac{1}{n_s - 1} \mathbf{X}^T \mathbf{X} \quad (3)$$

with n_s being the number of samples (rows) of the data matrix \mathbf{X} . The variance of the projection of \mathbf{X} onto a principal axis is

$$\text{var}\{\mathbf{X}\mathbf{w}_i\} = \mathbf{w}_i^T \mathbf{C} \mathbf{w}_i. \quad (4)$$

The covariance matrix is symmetric and therefore diagonal in the basis formed by its eigenvectors. Hence, using Eq. (4), one can see that the eigenvalues of \mathbf{C} are simply the variance of the respective principal component.^{25,28} For the actual implementation of PCA, we use the OpenCV library.²⁹

III. RESULTS

We start the results section analyzing the excluded area of regular polygons. Next, we show the PCA of the excluded area of randomly generated particles. We end the section with Monte Carlo simulations of selected particle shapes.

A. Regular polygons

The normalized values of the excluded areas for a line segment and for the regular polygons with 3–6 vertices are presented in Fig. 2. To check the accuracy of the numerical calculation of the excluded area, we have compared the numerical results against analytic expressions for line segments, equilateral triangles, and squares; see Figs. 2(b) and 2(c).

For regular polygons, the excluded area contains as many minima (and maxima) as vertices of the polygon since the rotational symmetry of the particle is also present in the excluded area. The excluded area of the line segment is a direct extension of this trend as it contains two minima (and two maxima).

All the excluded areas are symmetric with respect to the relative orientation $\phi = \pi$. This is a general property of the excluded area between any two identical particles that can be easily understood as follows. The relative orientations $-\phi$ and ϕ are degenerated as they correspond to an (irrelevant) swap of the two identical particles. Hence, the excluded area is symmetric with respect to $\phi = 0$, which implies also the symmetry with respect to $\phi = \pi$. Therefore, although in Fig. 2 we present the excluded area of particles with mirror symmetry, the symmetry of the excluded area around $\phi = \pi$ is also present for particles with no spatial symmetries.

As expected, the difference between the maximum value and the minimum value of the excluded area of regular polygons decreases by increasing the number of vertices of the regular polygon. The limit of a regular polygon with an infinite number of vertices is a disk, for which the excluded area does not depend on the relative orientation. The difference in the excluded area for different orientations is correlated with the increase in configurational entropy that occurs when the particles organize in a state with orientational order. Hence, ordering effects are stronger for particles with more variance in $A_{\text{exc}}(\phi)$. For example, in Ref. 30, Anderson *et al.* found that regular polygons with more than 6 vertices already melt like disks.

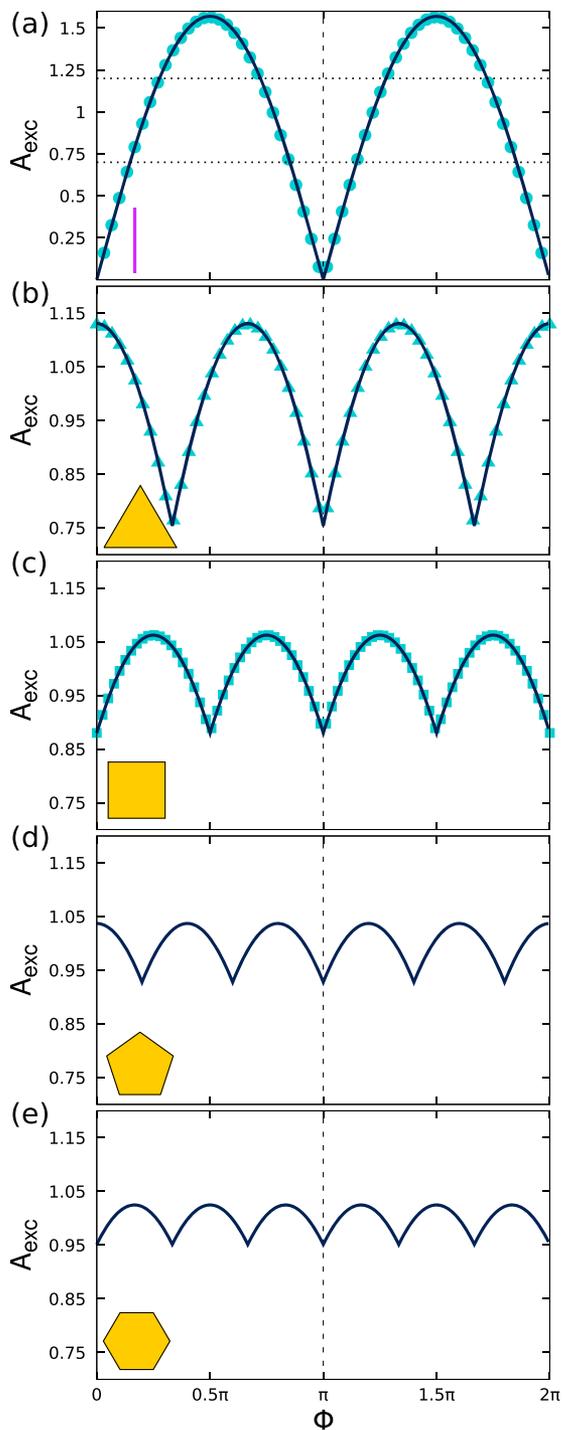


FIG. 2. Normalized excluded area for regular polygons as a function of the relative orientation ϕ : (a) line segment, (b) equilateral triangle, (c) square, (d) pentagon, and (e) hexagon. The insets show the particles. Solid lines are numerical results. The symbols are analytic expressions of the excluded area of a line segment (a), an equilateral triangle (b), and a square (c). The dotted horizontal lines in (a) indicate the vertical range used in panels (b)–(e).

B. Principal component analysis

In what follows, we show the result of the PCA applied to a set of 9.3×10^4 randomly generated particles of which 4.8×10^4 are polygons with a number of vertices uniformly distributed between 3 and 10, and 4.5×10^4 are star lines. An example of a non-self-intersecting polygon with 10 vertices is shown in Fig. 1(b). The number of vertices of the star lines is uniformly distributed between 2 and 10. We investigated 10 different sets, each one containing the same number of randomly generated particles. All sets produced the same results up to small numerical inaccuracies.

As described in Sec. II C, the first step in PCA consists of centering the data by removing the column-wise mean of the data matrix. The mean excluded area $m(\phi)$ [see Fig. 3(a)] has a global minimum at $\phi = \pi$ and a local minimum at $\phi = 0$. The mean excluded area manifests a common feature of all excluded areas we have calculated. The excluded area between two identical particles has always the global minimum at $\phi = \pi$, regardless of the shape of the particles. This intuitive feature has been proven for convex shapes.³¹ Our particles are both convex and concave, and we have always observed the global minimum to be located at $\phi = \pi$.

Panels (b)–(d) of Fig. 3 show the first three principal axes $w_i(\phi)$, $i = 1, 2, 3$. The first (b) and third (d) principal axes have qualitatively similar shapes to the excluded area of a line segment [Fig. 2(a)] and of an equilateral triangle [Fig. 2(b)], respectively. The second principal axis $w_2(\phi)$ [Fig. 3(c)] has a pronounced minimum at $\phi = \pi$. Excluded areas with this feature play an important role in the PCA analysis, as we discuss below. A data file with the mean excluded area and the first three principal components can be found in the [supplementary material](#).

In Fig. 4, a semilogarithmic plot of the eigenvalues of the first 10^2 principal axes is presented. The first and the second eigenvalue differ in one order of magnitude, and higher order eigenvalues decrease very fast in magnitude. Given that (i) the eigenvalue of a principal axis is the variance of the respective principal component^{25,28} and that (ii) the eigenvalues decrease very fast in magnitude (Fig. 4), we achieve a good representation of the excluded area using only the first three principal components. A measure of how well the data are represented using the first n components is the explained variance σ_n , which is the sum of the eigenvalues associated with the n first principal components divided by the sum of all eigenvalues. In our case, using three principal components, we find $\sigma_3 \approx 0.98$ and therefore we are confident that most of the full information is already contained in the first three components. Using the principal components of a given particle, its approximated excluded area can be reconstructed by calculating the sum of the first three principal axes [see Figs. 3(b)–3(d)] multiplied by their respective principal components and adding the mean [Fig. 3(a)]. In other words, the reconstructed excluded area $A_{\text{exc}}^{\text{rec}}$ is a linear combination of the principal axes and the mean (see Fig. 3), with the principal components being the coefficients of the linear combination

$$A_{\text{exc}}^{\text{rec}}(\phi) = m(\phi) + \sum_{i=1}^3 c_i w_i(\phi). \quad (5)$$

The average L_1 error

$$s_{L_1} = \langle |A_{\text{exc}}(\phi) - A_{\text{exc}}^{\text{rec}}(\phi)| \rangle \quad (6)$$

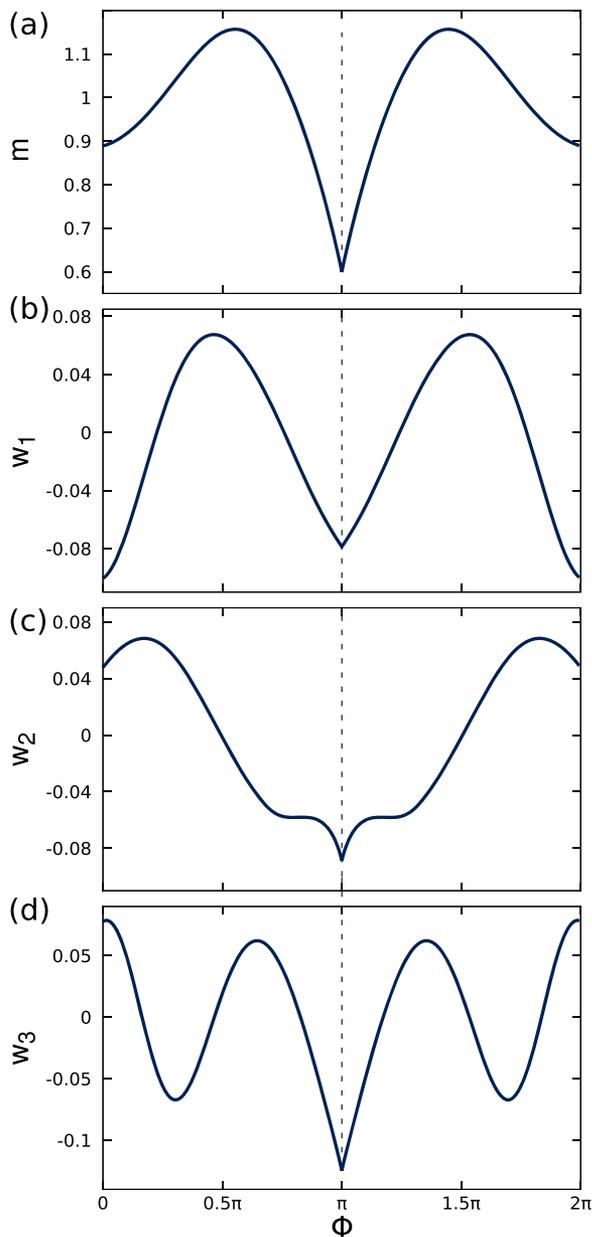


FIG. 3. Mean excluded area m (a), first w_1 (b), second w_2 (c), and third w_3 (d) principal axes as a function of the relative orientation between particles ϕ .

between the calculated values of the excluded areas $A_{\text{exc}}(\phi)$ and their reconstructions $A_{\text{exc}}^{\text{rec}}(\phi)$ is 0.012. Here, the average is taken over all orientations ϕ and over all samples. We show examples of the reconstructed excluded area for relevant particle shapes below.

In Fig. 5, we present the first three principal components c_i , $i = 1, 2, 3$ of the excluded area of all particles in the set. Each point represents the excluded area $A_{\text{exc}}(\phi)$ of one particle of the set.

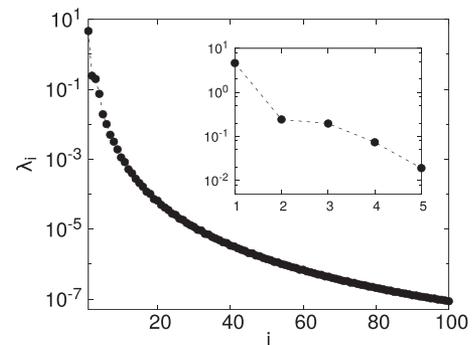


FIG. 4. Semilog plot of the first 10^2 eigenvalues λ_i vs the index i . The inset is a close view of the first five eigenvalues.

We also show two-dimensional projections onto the c_1 - c_2 , c_1 - c_3 , and c_2 - c_3 planes. The excluded areas are contained in a simply connected region with no holes and a well-defined boundary with four prominent limiting cases (highlighted by red squares in Fig. 5). We discuss in what follows these special limiting excluded areas found with PCA, together with their associated particle shapes.

1. First limiting case

As indicated by the eigenvalues, the first principal component c_1 has the by far highest variance with values between -3 and 7.5 . One vertex of the 3D projection of the excluded area [Fig. 5] corresponds to the line segment, which is the limiting case for which c_1 is maximized. In general, the value of the c_1 component increases with the elongation of the particles. According to PCA, the elongation of a particle is therefore the most important geometric feature influencing the excluded area. The excluded areas of the particles near this limiting case possess two well-defined minima, like in the case of a line segment shown in Fig. 2(a).

2. Second limiting case

Maximizing the second principal component c_2 is another limiting case of the 3D projection; see Fig. 5. An illustrative excluded area and its corresponding particle shape are presented in Fig. 6. The excluded areas in this region are characterized by a pronounced global minimum located at $\phi = \pi$ and a global maximum near $\phi = 0$. In Fig. 6, both the actual excluded area and the reconstruction using only the first three principal components are shown. The real excluded area has secondary minima that are not captured by the reconstructed excluded area. However, the overall agreement is very good and it justifies the use of only three principal components. The secondary features that are not reproduced by the reconstructed excluded area might play a role in the determination of the structure of phases with positional order, but it is less likely that they will affect the relative stability of fluid mesophases with only orientational order.

The particles in this region of the 3D projection are line stars with three arms. It might be possible to eliminate the secondary minima using shapes with curved lines (similar to the symbol ϵ). The shape of the excluded area $A_{\text{exc}}(\phi)$ suggests that the particles prefer a state where the neighboring particles are antiparallel.

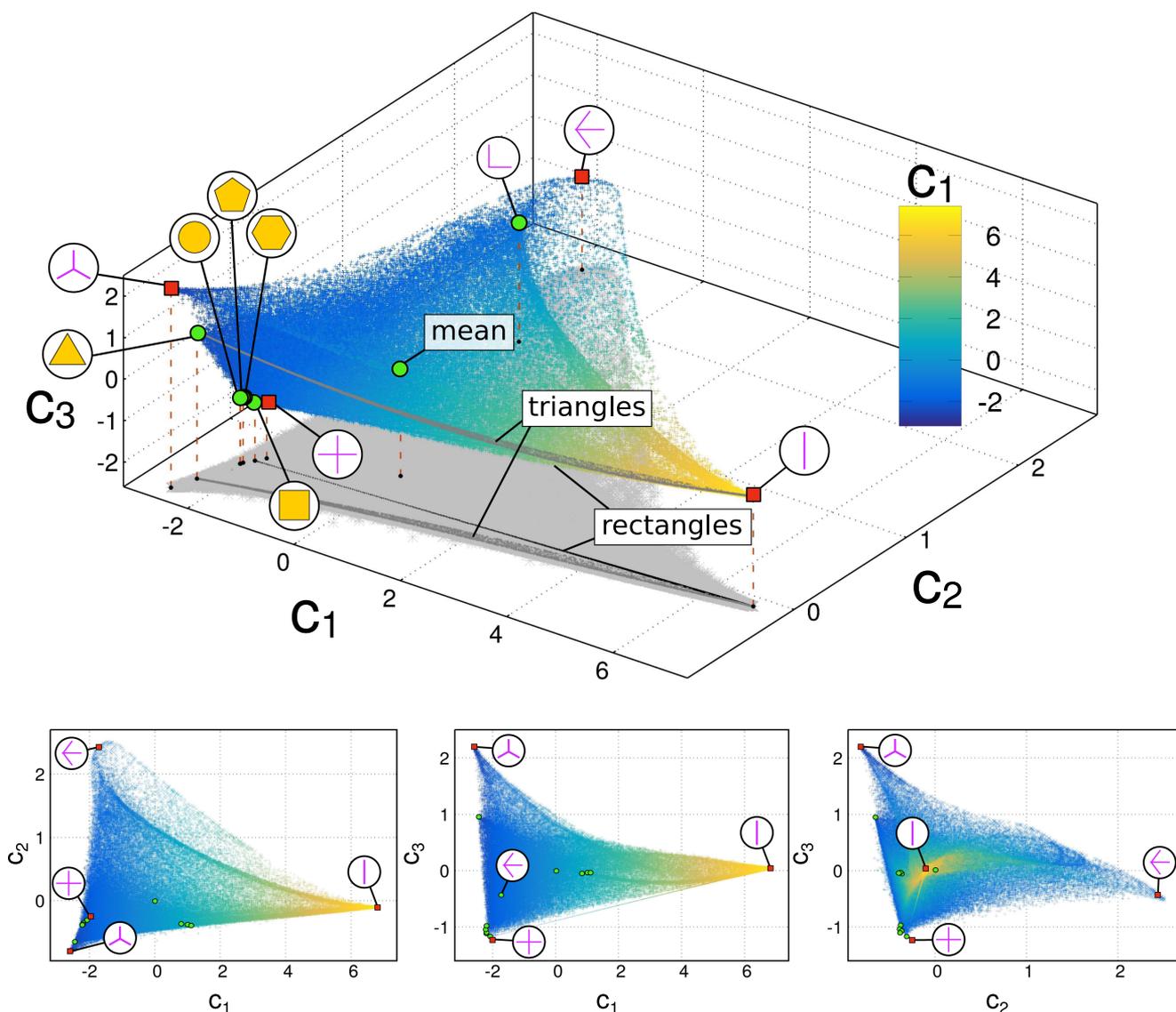


FIG. 5. Excluded areas in the space of the first three principal components. Top: The set of all values of the excluded area $A_{\text{exc}}(\phi)$ represented with the first three principal components c_i , $i = 1, 2, 3$. Each point represents one excluded area of the data set and is therefore associated to one particle shape. The color code indicates the value of the first principal component c_1 . Green circles show the values of the excluded areas of selected particles, as indicated. Red squares indicate the four limiting cases analyzed in the text. In light gray, a projection onto the c_1 - c_2 plane is shown. In dark gray, a projection of the excluded areas of triangles and rectangles is shown. Bottom: Two dimensional projections of the excluded areas onto the c_1 - c_2 (left), c_1 - c_3 (middle), and c_2 - c_3 planes.

A fluid of three-dimensional hard boomeranglike particles can form a biaxial nematic phase.³² In two dimensions, boomeranglike particles can be created by a star line with two arms at an angle of approximately 90° . In the space of the first three principal components, these particles are located relatively close to the second limiting case (see Fig. 5). The excluded area (not shown) has a prominent minimum at $\phi = \pi$, a local minimum if the particles are parallel $\phi = 0$, and depending on the relative angle between the two arms local minima at $\phi \approx \pi/2$ and $3\pi/2$. The presence of these local minima could stabilize a tetragonal phase.

3. Third limiting case

For the third limiting case emphasized by PCA, both c_1 and c_2 are minimal. The excluded areas located in this region have three pronounced equidistant minima. Before further discussing this case, we first have a look at the excluded areas for polygons with three vertices, i.e., triangles, which are closely related.

The excluded areas of triangles in the base of the first three principal axes are highlighted in dark gray in Fig. 5 and also represented in Fig. 7(a). All the excluded areas of triangles form a

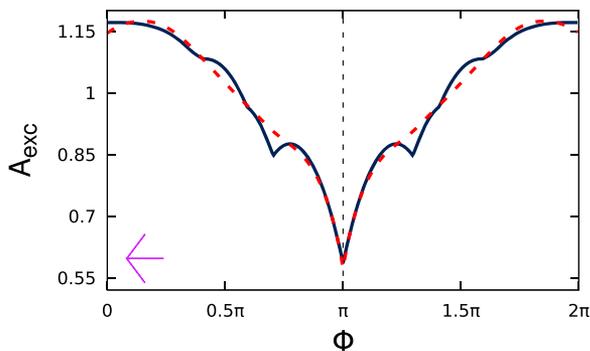


FIG. 6. Excluded area (solid black line) and reconstructed excluded area using the first three principal components (dashed red line) as a function of the relative orientation ϕ for the particle presented in magenta (left bottom corner).

simply connected two-dimensional region which spans between the two limiting cases given by a line segment and an equilateral triangle. These limiting cases correspond to triangles with the maximum (line segment) and minimum (equilateral triangle) possible aspect ratios. Going from the equilateral triangle to the line segment implies increasing the aspect ratio of a regular triangle. There are two special ways of elongating an equilateral triangle: (i) taking one side and

moving it away from the opposite vertex, making the angle at the apex in the resulting acute isosceles triangle very small, and (ii) taking one side and moving its two corresponding vertices away from each other along the direction of this side, creating an obtuse isosceles triangle where the angle at the apex is very large. In both cases, the intermediate triangles are isosceles and their excluded areas form the boundary of the 3D projection in the base of principal components; see Fig. 7(a). We have shown in Fig. 2(b) the excluded area of an equilateral triangle, which has three global minima at $\{\pi/3, \pi, 5\pi/3\}$ and three global maxima. In Fig. 7(b), we present the excluded area of other representative triangles. In panel (b1), we show an obtuse isosceles triangle. The excluded area has only one global minimum at $\phi = \pi$. The global minima of A_{exc} at $\pi/3$ and $5\pi/3$ of an equilateral triangle are now local minima and have moved to a different relative orientation. In panel (b2), we represent the excluded area of an acute isosceles triangle. There is only one global minimum at $\phi = \pi$, and the position of the secondary minima is also shifted with respect to that in an equilateral triangle. In panel (b3), we present a non-isosceles triangle, which in the PCA analysis is located between the two previous cases (b1) and (b2); see Fig. 7(a). The excluded area has characteristics of both acute and obtuse triangles. In all three cases, the reconstructions of the excluded areas neglect small features like the presence of local minima and kinks, but the overall agreement between the actual excluded area and that obtained with only three principal components is excellent.

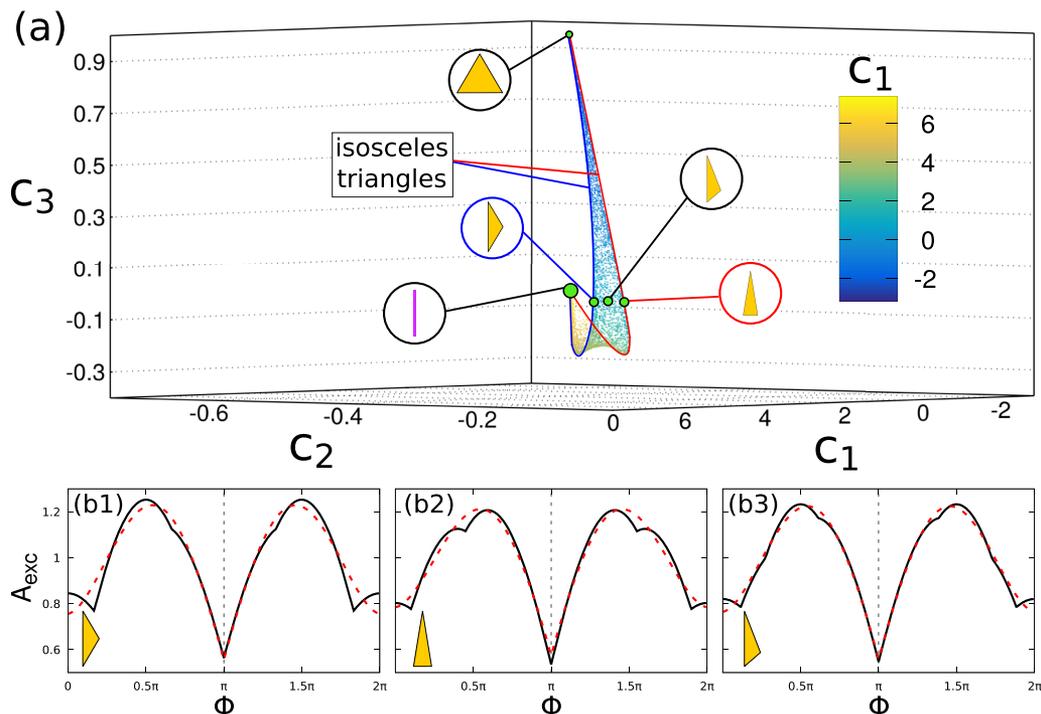


FIG. 7. (a) Excluded areas of triangles represented with the first three principal components, c_1 , c_2 , and c_3 . Each point represents the excluded area of one randomly generated triangle. The color indicates the value of the first principal component c_1 . Green circles indicate the position in the space of principal components of the excluded areas corresponding to the depicted particle shapes. The blue (red) curve indicates the location of all obtuse (acute) isosceles triangles. Bottom panels: Excluded areas (solid black line) and reconstructions using the first three principal components (dashed red line) as a function of the relative orientation ϕ for three particular triangles (represented in the left bottom corner of each panel).

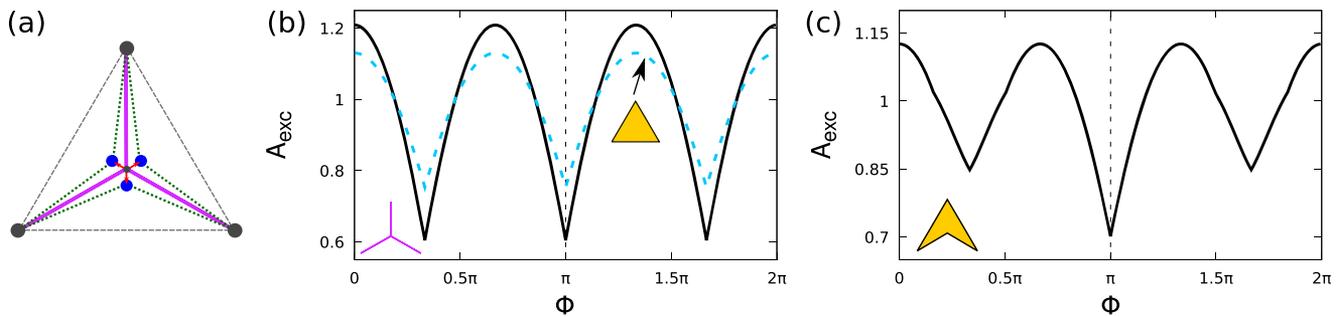


FIG. 8. (a) Schematic of the third limiting case highlighted by PCA: a star line with three identical segments rotated by $2\pi/3$ (magenta) such that the outer vertices lie on top of an equilateral triangle (dashed gray). Splitting the inner vertex into three vertices (blue dots) creates a new particle shape (dotted green). (b) Excluded area for a star line with three identical segments (solid black) and an equilateral triangle (dashed blue) as a function of the relative orientation ϕ . The star line is represented in magenta (bottom left corner). (c) Excluded area of an arrowlike particle, as shown in the bottom left corner.

In the space of principal components (see Fig. 5), the excluded area of an equilateral triangle is located in the region where both c_1 and c_2 are low, near the third limiting case for which c_1 and c_2 are minimal and c_3 is maximal. At this location, we find the excluded area of a star line particle made of three identical segments, any two of them forming an angle of $2\pi/3$. A sketch of the particle is presented in Fig. 8(a) together with the corresponding excluded area [Fig. 8(b)]. The excluded area resembles that of an equilateral triangle (also shown in the figure for comparison). Due to the normalization [Eq. (1)], the excluded area of the star line appears to have a larger variance (difference between the maximum and the minimum values of the excluded area). However, when normalizing the excluded area with the height of the particles, both excluded areas have the same variance. Strong differences are expected for three-body interactions and higher order terms. Hence, a comparison between the bulk phase behavior of this particle shape and that of equilateral triangles might help to understand the role of higher than two-body correlations on the bulk phase behavior.

Another interesting property of this kind of particles is that the star line can be continuously deformed by splitting its center and moving the resulting vertices radially toward the sides [see Fig. 8(a)]. As a special case, this includes particles where two of the inner vertices are located on top of the connecting line between two of the outer vertices. The resulting particles have four vertices and the shape of an arrow. The excluded area of such arrowlike particles is shown in Fig. 8(c). The main difference compared to the undeformed star particle is that the depth of two of the minima has decreased. However, the minima are still located at the same orientations as in the initial star line. This is in contrast to the case of triangles discussed above, for which any deformation simultaneously changes the depth of the two secondary minima as well as the relative orientations at which they occur. In this case, there is a complete family of particles in which the depth of the secondary minima can be tuned while keeping their location (relative orientation) fixed. Particles of this type could present an isotropic-triatic transition by increasing the density and then a second transition toward a uniaxial state at even higher densities.

4. Fourth limiting case

The last limiting excluded area according to PCA (see Fig. 5) occurs when c_3 is minimal. In this region, a number of special particles are located. Among them, we find disks, which have a completely flat excluded area independent of the relative orientation. Regular pentagons and hexagons together with other regular polygons with more vertices are located near the disk. This can be explained with the decreasing variance of the excluded area of regular polygons by increasing the number of sides, cf. Fig. 2. For example, the difference between the maximum and the minimum value of the excluded area for a line segment is approximately 1.57, whereas for the hexagon, it is approximately 0.07. It is therefore not surprising that extra principal components are required to have a good representation of the excluded areas in this region. As an example, we show in Fig. 9 the excluded area of pentagons and hexagons together with their reconstructed excluded areas. Up to six principal components are required to have a good approximation of the excluded area of hexagons.

The regular polygon with the lowest value of c_3 is the square. In the space of principal components [Fig. 5], there is a continuous curve containing all possible rectangles. At the end points of this curve, we find the square and the line segment, which are the rectangles with the minimal and maximal length-to-width aspect ratios, respectively.

We have shown above two particle shapes, an equilateral triangle and a star line with three identical segments that share very similar excluded areas; see Fig. 8(a). A similar behavior occurs for the case of a square and a star line with four identical segments forming the shape of a plus; see Fig. 10. The excluded area of the plus particle resembles that of the square but with higher variance (see a comparison in Fig. 10). The observed trend of decreasing variance in the excluded area for regular polygons as the number of sides increases holds also for the case of regular star lines.

The particle shapes of the four limiting excluded areas we have discussed above are well defined. This, however, is not the case for most excluded areas since different particle shapes can give rise to the same or almost the same excluded area. That two different hard bodies can produce the same excluded volume has been recently proven for convex bodies.³³

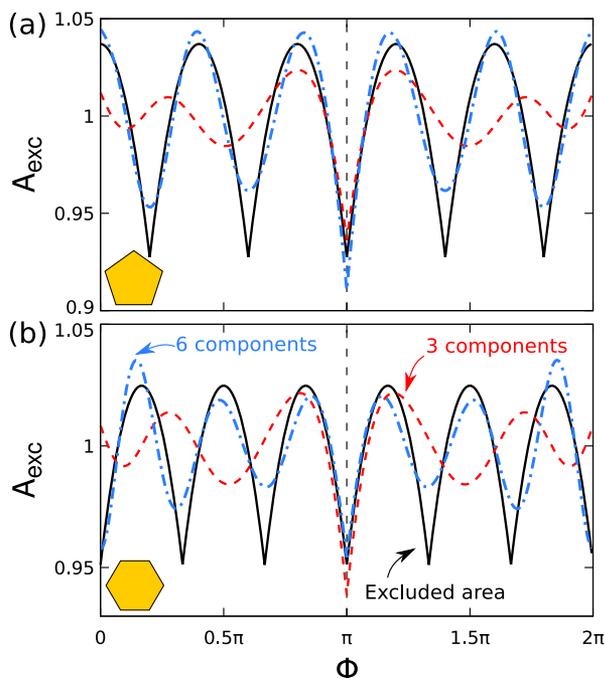


FIG. 9. Excluded area (black solid line) and reconstructed excluded area using three (red dashed line) and six (blue dotted line) principal components as a function of the relative orientation ϕ for the case of regular pentagons (a) and regular hexagons (b).

C. Monte Carlo simulations

We have shown above how PCA is useful to characterize the excluded areas of hard bodies (which play the role of the pair interaction potential in soft systems). The next natural step is a complete analysis of the bulk phase behavior of those particle shapes highlighted by PCA using, e.g., computer simulations or density functional theory. Some of the relevant shapes we have found, like rectangles^{7–9} and triangles,^{12,13} have been extensively studied, while others like the star lines have, to the best of our knowledge, not been

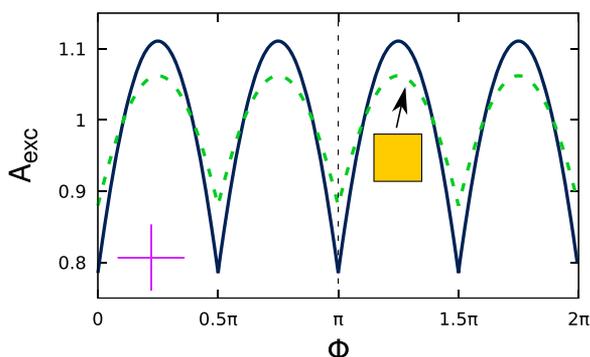


FIG. 10. Excluded area for the fourth limiting case highlighted by PCA, a star line with the shape of a plus (solid black line) and excluded area of a square (dashed green line) as a function of the relative orientation ϕ . The star line particle is shown in the bottom left corner.

analyzed yet. Although such analysis is not the scope of this work, we have also performed short Monte Carlo simulations for selected particle shapes in order to have an initial understanding of how the particle shape affects the bulk behavior.

We performed Monte Carlo (MC) simulations in the NpT (isothermal-isobaric) ensemble. We set a very high value of the pressure such that only steps that decrease the volume are accepted in order to compress the system from the ideal gas limit to very high packing fractions. In each simulation, a system of $N = 200$ particles is compressed at constant pressure. The particles are randomly initialized at sufficiently low density so that the stable state is isotropic. Then $\sim 5 \times 10^7$ Monte Carlo sweeps (MCS) are performed. Here, one MCS is an attempt to individually move and rotate all particles in the system. After every 1.5×10^4 MCS, an attempt to slightly change the volume of the system is performed. To this end, all particle positions are scaled accordingly. The maximum translation and rotation that each particle is allowed to perform in one MCS as well as the maximum volume change in one step are chosen such that the total acceptance probability is approximately 0.25. To detect if two particles overlap, we simply check for the intersection between all pairs of edges.⁵

We present here simulations for (i) particles that resemble an inverted umbrella [result of a deformation of a star line with three segments as indicated in Fig. 8(a)], (ii) arrow particles, and (iii) the polar particles we showed in Fig. 6. The shapes of the selected particles together with their excluded areas are shown in panel (a) of Fig. 11. Representative snapshots of particle configurations at high density are presented in panels (b) and (c). The particles ($i = 1, \dots, N$) are colored according to their $q_2^{(i)}$ (b) and $q_6^{(i)}$ (c) orientational order parameters, defined as

$$q_k^{(i)} = \left| \frac{1}{N_i} \sum_{j=1}^{N_i} e^{-ik\theta_j} \right|, \quad k = \{2, 6\}, \quad (7)$$

where θ_j is the orientation of particle j with respect to a fixed and arbitrary axis, and the sum runs over the N_i particles located at a distance smaller than approximately two particle lengths from particle i (including the i th particle). Note that the order parameter does not change if the relative orientation between the particles θ_{ij} is used in Eq. (7) instead of the absolute orientation of the particle θ_j (provided that the sum includes the i th particle).

The orientational order of the inverted umbrella particles is triatic (three directors forming an angle of $\pi/3$ between them), as indicated by the low q_2 [Fig. 11(b) left] and high q_6 [Fig. 11(c) left] values. By contrast, the orientational order of the arrowlike particles is uniaxial (particles oriented on average along one direction), with high values of both q_2 [Fig. 11(b) middle] and q_6 [Fig. 11(c) middle]. The different behavior between these two particle shapes can be explained with the properties of the excluded area. The excluded areas of both particles have the global minimum at $\phi = \pi$ and two local minima at $\phi = \pi/3$ and $5\pi/3$; see Fig. 11(a). The difference lies in the ratio between the depths of the local minima and the global minimum (measured from the global maximum), which is 0.77 for the inverted umbrella [Fig. 11(a) left] and 0.65 for the arrowlike particle [Fig. 11(a) middle].

The relative stability of phases with triatic and uniaxial orientational order is the result of a competition between orientational and configurational entropies. In a triatic phase, the particles populate

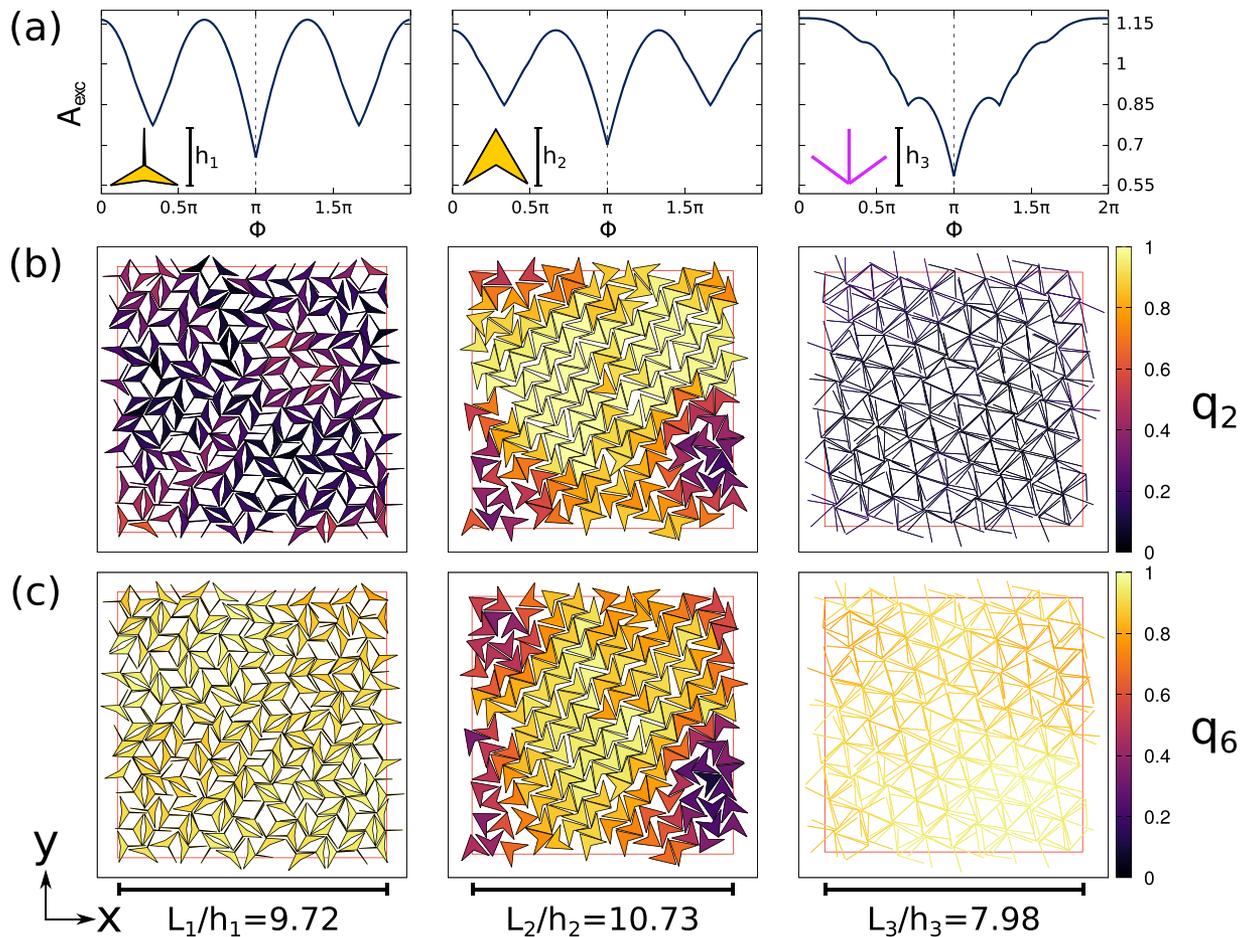


FIG. 11. (a) Excluded areas and geometry of three selected particles. Panels (b) and (c) show selected snapshots of a Monte Carlo simulation of $N = 200$ particles. The color code of the particles indicates the local particle-based order parameters q_2 (b) and q_6 (c). The order parameters of each particle are computed taking into account all particles located in a circular region of diameter two particle lengths centered at the particle of interest.

the three possible minima of the excluded area and hence the orientational entropy is higher than that in the uniaxial phase. However, in a uniaxial phase, only the global minimum of the excluded area is populated (for which the packing is more efficient), which results in higher configurational entropy compared to the triatic phase. The result of this competition depends on the relative depth between the local and the global minima. This parameter can be continuously tuned via deformation of the particle shape.

Note also that there is no minimum at $\phi = 0$ for the excluded area of the arrow particles. Therefore, in the uniaxial phase of the arrow particles, neighboring particles always point in opposite directions. Our simulations suggest a possible coexistence between this ordered state and an isotropic state for the arrow particles. Note that a small region of the simulation box remains isotropic. However, further longer and larger simulations are required to study this in detail.

For the third type of particles [Fig. 11(a) right], the excluded area has a well pronounced minimum at $\phi = \pi$. According to this

observation, one could expect a uniaxial state similar to that found for the arrow particles. The simulations, however, reveal that the particles prefer a state with triatic order. It is then clear that higher than two-body correlations are required to predict the bulk phase behavior for this particle shape. This resembles the case of hard rectangles. There, particles with length-to-width aspect ratio smaller than ~ 7 form small clusters of a few particles each. The cluster formation leads to a global tetraic order^{34–36} with two perpendicular directors in which the symmetry differs from that of the particle shape.

IV. DISCUSSION AND CONCLUSIONS

We have used principal component analysis to investigate the excluded area of two-dimensional hard-bodies. We have restricted the analysis to particle shapes given as non-self-intersecting polygons and star lines. In both cases, the particle shapes were randomly generated. Our analysis shows that the magnitude of the

excluded area as a function of the relative particle orientation is dominated by global features. Hence, despite the vast diversity of particle shapes, the variety of possible excluded areas is more restricted. A clear indication of this is that only three principal components are required to produce a good approximation of the magnitude of the excluded area as a function of the relative particle orientation.

One feature shared by all excluded areas is the position of the global minimum, which in our $\sim 10^6$ randomly generated samples always occurs when the two particles are antiparallel, forming a relative angle $\phi = \pi$. This result, known for convex bodies,³¹ seems to also hold for nonconvex particles.

The dimension reduction via PCA identifies the elongation of the particles as the most prominent feature affecting the excluded area. Furthermore, when represented with the first three principal components, all excluded areas are located on a simply connected three-dimensional region with no holes. At the boundaries, we find four well-defined limiting cases corresponding to shapes of the excluded area with one to four equidistant prominent minima. The relative depth between these prominent minima is a parameter that can be adjusted by varying the particle shape. In contrast to this flexibility, there is not much freedom to vary the relative angle at which the prominent minima of the excluded area occur. However, other features of the excluded area such as the position and number of secondary minima can be tuned. While we do not expect a high impact on the fluid phases, the secondary features might play an important role on the stability of phases with positional order.

We have identified different particle shapes with very similar excluded areas. These particles are ideal candidates to investigate the role of higher than two-body correlations on the bulk phase behavior. Higher than two-body correlations can even dominate the behavior of the system, in which case Onsagerlike theoretical approaches³⁷ that rely entirely on two-body correlations might fail to describe the phenomenology of the system. We have shown an example using MC simulation in which the excluded area possesses a unique global minimum at $\phi = \pi$, but the orientational order of the particles is triatic.

The phase behavior of three particle shapes highlighted in our PCA analysis has been studied: line segments form uniaxial nematic phases,⁶ a fluid of squares¹⁴ or rectangles with short aspect ratio form a tetratic phase,⁷ and triangles form triatic phases.¹² That is, the particle shapes highlighted in the PCA analysis give rise to the formation of mesophases with different orientational properties. PCA is therefore a powerful technique to classify the interaction in hard models and to anticipate the particle shapes of potential interest. PCA can complement other approaches intended to understand self-assembly in colloidal systems such as the inverse design of pair potentials³⁸ and the systematic study of regular shapes using computer simulations.^{39,40} Recently it has been shown how PCA can be applied to detect phase transitions in lattice⁴¹ as well as in continuous⁴² systems.

Promising extensions of the current work are the application of PCA to the excluded area of three-dimensional hard bodies and binary mixtures. Regarding binary mixtures, we expect a much richer variety of shapes. Note, for example, that the excluded area between two different particles does not have to be symmetric with respect to a certain relative orientation (in contrast to the excluded

area between identical particles, which is always symmetric with respect to $\phi = \pi$).

SUPPLEMENTARY MATERIAL

See [supplementary material](#) for a data file with the principal components.

ACKNOWLEDGMENTS

We thank M. Schmidt, E. Velasco, and Y. Martínez-Ratón for feedback and stimulating discussions.

REFERENCES

- 1 B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* **27**, 1208 (1957).
- 2 L. Mederos, E. Velasco, and Y. Martínez-Ratón, *J. Phys.: Condens. Matter* **26**, 463101 (2014).
- 3 Y. Rosenfeld, M. Schmidt, H. Löwen, and P. Tarazona, *Phys. Rev. E* **55**, 4245 (1997).
- 4 A. L. Thornework, J. L. Abbott, D. G. A. L. Aarts, and R. P. A. Dullens, *Phys. Rev. Lett.* **118**, 158001 (2017).
- 5 R. Roth, K. Mecke, and M. Oettel, *J. Chem. Phys.* **136**, 081101 (2012).
- 6 D. Frenkel and R. Eppenga, *Phys. Rev. A* **31**, 1776 (1985).
- 7 Y. Martínez-Ratón, E. Velasco, and L. Mederos, *J. Chem. Phys.* **122**, 064903 (2005).
- 8 A. Donev, J. Burton, F. H. Stillinger, and S. Torquato, *Phys. Rev. B* **73**, 054109 (2006).
- 9 T. Geigenfeind, S. Rosenzweig, M. Schmidt, and D. de las Heras, *J. Chem. Phys.* **142**, 174701 (2015).
- 10 M. A. Bates and D. Frenkel, *J. Chem. Phys.* **112**, 10034 (2000).
- 11 J. Dzubiella, M. Schmidt, and H. Löwen, *Phys. Rev. E* **62**, 5081 (2000).
- 12 A. P. Gantapara, W. Qi, and M. Dijkstra, *Soft Matter* **11**, 8684 (2015).
- 13 Y. Martínez-Ratón, A. Díaz-De Armas, and E. Velasco, *Phys. Rev. E* **97**, 052703 (2018).
- 14 K. Wojciechowski and D. Frenkel, *Comput. Methods Sci. Technol.* **10**, 235 (2004).
- 15 C. Avendaño and F. A. Escobedo, *Soft Matter* **8**, 4675 (2012).
- 16 T. Schilling, S. Pronk, B. Mulder, and D. Frenkel, *Phys. Rev. E* **71**, 036138 (2005).
- 17 W. van der Stam, A. P. Gantapara, Q. A. Akkerman, G. Soligno, J. D. Meeldijk, R. van Roij, M. Dijkstra, and C. de Mello Donega, *Nano Lett.* **14**, 1032 (2014).
- 18 Z. Hou, K. Zhao, Y. Zong, and T. G. Mason, *Phys. Rev. Mater.* **3**, 015601 (2019).
- 19 J. A. Cuesta and D. Frenkel, *Phys. Rev. A* **42**, 2126 (1990).
- 20 S. Varga, P. Gurin, J. C. Armas-Perez, and J. Quintana-H, *J. Chem. Phys.* **131**, 184901 (2009).
- 21 J. A. Martínez-González, S. Varga, P. Gurin, and J. Quintana-H, *J. Mol. Liq.* **185**, 26 (2013).
- 22 J. A. Martínez-González, S. Varga, P. Gurin, and J. Quintana-H, *Europhys. Lett.* **97**, 26004 (2012).
- 23 E. S. Harper, R. L. Marson, J. A. Anderson, G. Van Anders, and S. C. Glotzer, *Soft Matter* **11**, 7250 (2015).
- 24 K. Pearson, *Philos. Mag.* **2**, 559 (1901).
- 25 S. Wold, K. Esbensen, and P. Geladi, *Chemom. Intell. Lab. Syst.* **2**, 37 (1987).
- 26 G. A. Croes, *Oper. Res.* **6**, 791 (1958).
- 27 S. Lin, *Bell Syst. Tech. J.* **44**, 2245 (1965).
- 28 I. Jolliffe, *Principal Component Analysis* (Springer Verlag, New York, 2002).
- 29 G. Bradski, Dr. Dobb's J. Software Tools **25**, 120–125 (2000).
- 30 J. A. Anderson, J. Antonaglia, J. A. Millan, M. Engel, and S. C. Glotzer, *Phys. Rev. X* **7**, 021001 (2017).

- ³¹P. Palfy-Muhoray, E. G. Virga, and X. Zheng, *J. Phys. A: Math. Theor.* **47**, 415205 (2014).
- ³²P. I. C. Teixeira, A. J. Masters, and B. M. Mulder, *Mol. Cryst. Liq. Cryst. Sci. Technol., Sect. A* **323**, 167 (1998).
- ³³J. M. Taylor, *J. Phys. A: Math. Theor.* **52**, 095002 (2019).
- ³⁴Y. Martínez-Ratón and E. Velasco, *Phys. Rev. E* **79**, 011711 (2009).
- ³⁵Y. Martínez-Ratón, E. Velasco, and L. Mederos, *J. Chem. Phys.* **125**, 014501 (2006).
- ³⁶T. Müller, D. de las Heras, I. Rehberg, and K. Huang, *Phys. Rev. E* **91**, 062207 (2015).
- ³⁷L. Onsager, *Ann. N. Y. Acad. Sci.* **51**, 627 (1949).
- ³⁸C. S. Adorf, J. Antonaglia, J. Dshemuchadse, and S. C. Glotzer, *J. Chem. Phys.* **149**, 204102 (2018).
- ³⁹D. Klotsa, E. R. Chen, M. Engel, and S. C. Glotzer, *Soft Matter* **14**, 8692 (2018).
- ⁴⁰D. Wan, C. X. Du, G. van Anders, and S. C. Glotzer, preprint [arXiv:1901.09523](https://arxiv.org/abs/1901.09523) (2019).
- ⁴¹S. J. Wetzel, *Phys. Rev. E* **96**, 022140 (2017).
- ⁴²R. B. Jadrich, B. A. Lindquist, and T. M. Truskett, *J. Chem. Phys.* **149**, 194109 (2018).