Maximum probability domains for the analysis of the microscopic structure of liquids

Federica Agostini, Giovanni Ciccotti, Andreas Savin, and Rodolphe Vuilleumier

1 Max-Planck Institut für Mikrostrukturphysik, Weinberg 2, D-06120 Halle, Germany.
2 Dipartimento di Fisica and CNISM Unità 1, Università “La Sapienza”, Piazzale Aldo Moro 5, I-00185 Rome, Italy.
3 School of Physics, University College Dublin (UCD), Belfield Dublin 4, Ireland.
4 CNRS, UMR 7616, Laboratoire de Chimie Théorique, F-75005 Paris, France.
5 Sorbonne Universités, UPMC Univ Paris 06, UMR 7616, Laboratoire de Chimie Théorique, F-75005 Paris, France.
6 UMR 8640 ENS-CNRS-UPMC, Département de Chimie, 24 rue Lhomond, École Normale Supérieure, 75005 Paris, France.
7 UPMC Université Paris 06, 4, Place Jussieu, 75005 Paris, France.

Abstract

We introduce the concept of maximum probability domains, developed in the context of the analysis of electronic densities, in the study of the microscopic spatial structures of liquids. The idea of locating a particle in a three dimensional region, by determining the domain where the probability of finding that, and only that, particle is maximum, gives an interesting characterisation of the local structure of the liquid. The optimisation procedure, required for the search of the domain of maximum probability, is carried out by the implementation of the level set method. Some results for few case studies are presented. In particular by looking at liquid water at different densities or at the solvation shells of Na⁺ always in liquid water.
I. INTRODUCTION

While liquids have long-range disorder, they exhibit a local order that drives many of their physical properties. This short-range order is usually characterised through the Radial Distribution Functions (RDFs) \([1]\). However, the RDFs convey little or indirect information about the three-dimensional (3D) local structure of liquids. Spatial Distribution Functions (SDFs) have then been introduced for this purpose \([2, 3]\). Since then, several indicators of the local order of liquids have been proposed, in particular for the important case of liquid water, including the statistics of Voronoi polyhedron \([4, 8]\), local structure index \([9]\) or angular correlation \([10, 11]\). Recently, local density fluctuations and the probability distribution of occupancy of a given volume have attracted a lot of attention, in the framework of the theory of hydrophobicity \([12]\). These have been used to characterise and locate patches of hydrophobicity or hydrophilicity at surfaces \([13, 15]\) or around proteins \([13, 16]\). However, as such, they have not been used to characterise the microscopic structure of liquids.

In a very different context, that of electronic structure, related concepts have been used to locate electron pairs \([17-19]\). In general, these methods are designed to identify regions in 3D space with particular chemical and physical meaning. The analysis is based on the information provided by the electronic probability density calculated from the many-body wave function, but offers a way to define and to visualise some relevant regions in real space.

The work presented in this paper is motivated by the interest for developing an alternative approach to study the local structure of liquid systems, combined with the clear similarities between the electronic probability density and the atomic many-body probability distribution of a liquid. We propose here a method based on the identification of regions of space where the probability of finding one and only one particle is maximum. We refer to these regions as Maximum Probability Domains (MPDs). The definition of this probability is given as the generalisation of the concept \([17, 20, 23]\) used for the analysis of electronic probability densities in molecules and indeed can be generalised to the case of \(n\) (and only \(n\)) particles. The information encoded in this quantity has a many-body nature, also when considering the one-particle occupancy probability. It has to be viewed as the probability of finding \(n\) particles within a certain region of space with all other particles outside. Therefore, such probability is different from any reduced \(n\)-particle probability density, since the degrees of freedom of the \(N - n\) remaining particles (\(N\) being the total number of particles in the system) are not integrated out. In analogy to such quantities, however, the probability can be evaluated by using a standard sampling procedure with molecular dynamics (MD) or Monte Carlo (MC) trajectories.

We illustrate the method by applying it to describe the structure of pure liquid water and to define the solvation shells of sodium ions in a diluted water solution. We compare the results of our study with more standard approaches based on the RDFs and with various definitions of solvation shells. It will become apparent from the illustrations that the method offers an efficient complementary tool to the analysis of liquid structures, both qualitatively and quantitatively. The one-particle MPD defines the boundary of the 3D region available to a given particle, where it is not likely that another particle can penetrate. When a set of MPDs is identified around a central water molecule or sodium ion, we obtain a 3D-map of the locations of the surrounding molecules with a number of MPDs that is usually smaller, as we will see later, than the standard coordination number. The request of exclusive occupation of a MPD by only one particle is responsible for this feature, even though for rigid local structures, as the first solvation shell of water, there are clear similarities between the two pictures (MPDs and coordination numbers). In the case of sodium in water, the one-particle MPDs are located around the central ion such that the spherical symmetry of the problem is maintained. However, the shape and dimension of the domains give a 3D resolution that is not accessible with the SDFs.

The paper is organised as follows. First, we introduce the general definition of the one-particle occupancy probability. Then, the maximisation of such probability is posed as a geometric optimisation problem and solved within the framework of the level set method (LSM). We give the details for the numerical implementation of the method and we present the two illustrations mentioned before. Finally, we conclude and state our outlook for future developments.

II. DEFINITION OF MPDS

The one-particle MPD is defined as the region in 3D space where the probability of finding one, and only one, particle is maximum. In the present context, the term particle stands for an atom or the centre of mass of a molecule or some other point of high symmetry in a molecule. The choice depends on the information that we want to extract from this analysis. For instance, in the study of liquid water, as presented in this paper, the particles will be the oxygen atoms in the molecules.

If a system is composed of \(N + 1\) identical particles with positions \(\mathbf{R}'^{N+1} = \mathbf{R}_0', \mathbf{R}_1', \ldots, \mathbf{R}_N'\) and we focus on the \(N\) particles with positions \(\mathbf{R}_i = \mathbf{R}_i' - \mathbf{R}_0'\) for \(i = 1, \ldots, N\), we can define the one-particle occupancy probability of a domain \(\Delta\), \(P^{(1)}(\Delta)\), as the probability of finding one, and only one, particle in the region of space \(\Delta\) with all other
particles located outside $\Delta$. This probability \[21\], written as

$$P^{(1)}(\Delta) = \sum_{n=0}^{N} \int_{\Delta} dR_n \int_{\Delta_{c}} dR_n \int_{\Delta} dR_{n+1} \ldots dR_N \rho(R^N), \tag{1}$$

is what in mathematics is called a set function. Here, $\rho(R^N)$ is the configurational probability density and, in the particular case of water and sodium diluted in water, will be explicitly defined below in Eqs. \[8\] to \[10\].

The domain $\Delta$ is given in a reference frame that is centred on the particle at $R_0$, which is kept out of statistics (in the sense that in Eq. \[1\] we do not integrate over its positions). $\Delta_c$ is the symbol used to indicate the complementary volume of $\Delta$. The sum in Eq. \[1\] allows us to define the probability that is independent of the identity of the particle occupying the domain $\Delta$. Moreover, the notation in Eq. \[1\] indicates that all integrals involving the variables $R_1 \ldots R_{i-1}$ and $R_{i+1} \ldots R_N$ are performed over $\Delta_c$. This definition of the probability can be easily extended to $n$ particles in $\Delta$ and $N-n$ particles in $\Delta_c$. Eq. \[1\] satisfies the normalisation condition

$$\sum_{n=0}^{N} P^{(n)}(\Delta) = 1, \tag{2}$$

which can be easily obtained for an ideal gas (in the absence of interactions among particles) occupying a volume $V$, as the configurational probability density is explicitly known, i.e. $\rho(R^N) = V^{-N}$, thus allowing for the analytic calculation of Eq. \[1\]. This is done by inserting the expression of the one-particle occupancy probability $P^{(n)}(\Delta) = (\frac{v}{V})^n (V - v)^{N-n} / V^N$ in Eq. \[2\], where $v$ is the volume occupied by $\Delta$ and $V - v$ that occupied by $\Delta_c$.

To extend the integration domain in Eq. \[1\] to the whole configuration space, we introduce the characteristic function $\Upsilon_{\Delta}(R)$, defined as

$$\Upsilon_{\Delta}(R) = \begin{cases} 1 & \text{if } R \in \Delta \\ 0 & \text{otherwise} \end{cases}, \quad \Upsilon_{\Delta_c}(R) = 1 - \Upsilon_{\Delta}(R). \tag{3}$$

The probability $P^{(1)}(\Delta)$ is then

$$P^{(1)}(\Delta) = \sum_{i=1}^{N} \int dR_i \Upsilon_{\Delta}(R_i) \prod_{j \neq i}^{N} (1 - \Upsilon_{\Delta}(R_j)) \rho(R^N) \tag{4}$$

which allows us to identify the microscopic observable

$$\Gamma_{\Delta}^{(i)}(R^N) = \Upsilon_{\Delta}(R_i) \prod_{j \neq i}^{N} (1 - \Upsilon_{\Delta}(R_j)). \tag{5}$$

The equilibrium average of $\Gamma_{\Delta}^{(i)}(R^N)$, evaluated according to the probability density $\rho(R^N)$, leads to the definition of $P^{(1)}(\Delta)$

$$P^{(1)}(\Delta) = \left\langle \sum_{i=1}^{N} \Gamma_{\Delta}^{(i)}(R^N) \right\rangle. \tag{6}$$

Eq. \[6\] can be evaluated by sampling the microscopic observable $\Gamma_{\Delta}^{(i)}(R^N)$ along a MD or MC trajectory. From the algorithmic point of view, $P^{(1)}(\Delta)$ can be estimated by computing

$$\nu^{(1)}(\Delta) = \frac{1}{N_{\text{conf}}} \sum_{\kappa=1}^{N_{\text{conf}}} \delta_{1,n_{\Delta}(\kappa)} \simeq P^{(1)}(\Delta), \tag{7}$$

the frequency of the event “one, and only one, particle inside $\Delta$”. Here, $N_{\text{conf}}$ is the total number of configurations sampled along the trajectory and $\kappa$ labels the selected configuration. The term in the sum is equal to one only if the number of particles $n_{\Delta}(\kappa)$ inside the volume $\Delta$, for the configuration $\kappa$, is equal to 1.

We intend to illustrate how to use the MPDs to investigate the local structure of a liquid around a given centre. In the examples discussed later on, the centre will be chosen as a water molecule, in the case of liquid water, and as the sodium ion, in the case where we analyse the structure of the water solvation shells around this ion. Therefore,
the probability density used in Eq. [1] has to be identified as the conditional probability density of the configuration \( \mathbf{R}^N \) of the \( N \) oxygen atoms, given a certain configuration \( \{ \mathbf{R}_0, \mathbf{R}_0^{H_1}, \mathbf{R}_0^{H_2} \} \) of the central water molecule,

\[
\rho(\mathbf{R}^N) = \rho \left( \mathbf{R}^N | \mathbf{R}_0, \mathbf{R}_0^{H_1}, \mathbf{R}_0^{H_2} \right),
\]  

(8)

or of the central sodium ion \( \{ \mathbf{R}^N_{0Na^+} \} \),

\[
\rho(\mathbf{R}^N) = \rho \left( \mathbf{R}^N | \mathbf{R}_{0Na^+} \right).
\]

(9)

We explicitly discuss below only the case of water, while the expressions for sodium in water are obtained by replacing \( \{ \mathbf{R}^N_{0Na^+} \} \) to \( \{ \mathbf{R}_0, \mathbf{R}_0^{H_1}, \mathbf{R}_0^{H_2} \} \).

Here we label with "0" the positions of the oxygen atom and the two hydrogen atoms of this central molecule. To be more explicit on the relation between this density and the standard configurational canonical density \( \{ \mathbf{R}_0, \mathbf{R}_0^{H_1}, \mathbf{R}_0^{H_2}, \mathbf{R}_0^{H_1, N}, \mathbf{R}_0^{H_2, N} \} \), that is a function of all atomic positions, i.e. oxygens and hydrogens, we give the expression of \( \rho(\mathbf{R}^N) \) in terms of this density,

\[
\rho(\mathbf{R}^N) = \int d\mathbf{R}_0^{H_1, N} d\mathbf{R}_0^{H_2, N} \rho_{\text{can}} \left( \mathbf{R}_0, \mathbf{R}_0^{H_1, N}, \mathbf{R}_0^{H_2, N} \right) P_m \left( \mathbf{R}_0, \mathbf{R}_0^{H_1}, \mathbf{R}_0^{H_2} \right).
\]

(10)

Since \( \rho(\mathbf{R}^N) \) is a conditional probability density, the denominator represents the marginal probability of finding the central molecule in the configuration \( \{ \mathbf{R}_0, \mathbf{R}_0^{H_1}, \mathbf{R}_0^{H_2} \} \), namely

\[
P_m \left( \mathbf{R}_0, \mathbf{R}_0^{H_1}, \mathbf{R}_0^{H_2} \right) = \int d\mathbf{R}_0^{H_1, N} d\mathbf{R}_0^{H_2, N} \rho_{\text{can}} \left( \mathbf{R}_0, \mathbf{R}_0^{H_1, N}, \mathbf{R}_0^{H_2, N} \right).
\]

(11)

The construction of the MPDs is carried out by initially choosing some centres around the central particle (labeled by 0) and some volumes enclosing these centres. A set of MPDs will be then identified for each centre as

\[
\Delta_i^* = \arg \max_{\Delta_i} P^{(1)}(\Delta_i), \quad \text{with } i = 1, 2, \ldots
\]

(12)

The centres around the 0th particle can be chosen quite arbitrarily, however we find useful to locate them in correspondence of the local maxima of the SDF or of the RDF. It is worth underlining that \( \Delta_i^* \) is not an absolute maximum over all space of volumes, but it is the one found by a local search (the closest to the initial choice).

The definition of the set of MPDs \( \Delta_1^*, \Delta_2^*, \ldots \) can be used for the characterisation of the structure of atomic and molecular liquids as an alternative or a complement to standard tools such as distribution functions, solvation shells and coordination numbers. The determination of the optimised domains \( \Delta_i^* \)'s is posed as a geometric optimisation problem, since the probability \( P^{(1)}(\Delta_i) \) has to be maximised with respect to variations of \( \Delta_i \). In the following section we present \cite{[24–28]} the LSM to be applied in the search of the MPDs, in particular in a liquid.

III. APPLICATION OF THE LEVEL SET METHOD

According to the definition given in Eq. [1], \( P^{(1)}(\Delta) \) is a set function, since it depends on the set of points that determines the region \( \Delta \). To solve Eq. [12] and find the maximum of this set function, we apply the LSM, presented in Appendix A. There, we show that the evolution of the initial domain \( \Delta_i \) to get the maximising set, is given by

\[
\Delta \rightarrow \Delta_r = \{ \mathbf{R}_r = \mathbf{R} + \mathbf{v}(\mathbf{R})d\tau | \mathbf{R} \in \partial \Delta \},
\]

(13)

where \( \mathbf{v}(\mathbf{R}) \) is interpreted as a velocity field. Here, the deformation of the domain \( \Delta \) is expressed as a time evolution. However, \( d\tau \) is a fictitious time step, as well as \( \mathbf{v}(\mathbf{R}) \) is a fictitious velocity. In Appendix A we obtain that

\[
\mathbf{v}(\mathbf{R}) = \mathbf{n}(\mathbf{R}) f_\Delta(\mathbf{R})
\]

(14)

with \( \mathbf{n}(\mathbf{R}) \) the unitary vector normal to the boundary of \( \Delta \). Moreover, \( f_\Delta(\mathbf{R}) \), the modulus of the velocity field, is

\[
f_\Delta(\mathbf{R}) = \left\{ \sum_{i=1}^{N} \left[ \delta_{\mathbf{RR}_i} \prod_{j \neq i}^{N} (1 - \Upsilon_\Delta(\mathbf{R}_j)) - \Upsilon_\Delta(\mathbf{R}_i) \sum_{k \neq i}^{N} \delta_{\mathbf{RR}_k} \prod_{j \neq i,k}^{N} (1 - \Upsilon_\Delta(\mathbf{R}_j)) \right] \right\}.
\]

(15)
As discussed in the appendix, the level set function \( \phi \) function is any smooth function which can be defined by following the prescriptions:

\[
\Gamma_i^{(i)}(R; R^N) = \delta_{RR_i} \prod_{j \neq i}^{N} (1 - Y_{\Delta}(R_j)) - Y_{\Delta}(R_i) \sum_{k \neq i}^{N} \delta_{RR_k} \prod_{j \neq i,k}^{N} (1 - Y_{\Delta}(R_j)).
\] (16)

Therefore, the quantity given in Eq. (16) can be numerically sampled along a MD or MC trajectory. To evaluate \( \Gamma_i^{(i)}(R; R^N) \), we proceed as follows:

- First term on the right-hand-side of Eq. (16): if \( R_i = R \) and all the other \( N - 1 \) particles are in \( \Delta_c \), then the term is 1, irrespective of whether \( R \) is in \( \Delta \) or in \( \Delta_c \); otherwise it is 0.

- Second term on the right-hand-side of Eq. (16): if \( \exists k \neq i \) such that \( R_k = R \), while \( R_i \in \Delta \) and the other \( N - 2 \) \( R_j \in \Delta_c \), then the term is 1, irrespective of whether \( R \) is in \( \Delta \) or in \( \Delta_c \); otherwise it is 0.

As discussed in the appendix, the level set function \( \phi(R, \tau) \) can be used to track \( \Delta \) at each time step. The level set function is any smooth function which can be defined by following the prescriptions:

(i) \( \phi(R, \tau) = 0 \) \( \forall \tau \) if \( R_i \in \partial \Delta \);

(ii) the gradient of \( \phi(R, \tau) \) for \( R_i \in \partial \Delta \) has to be parallel to \( n(R_i) \);

(iii) the evolution of \( \phi(R, \tau) \) is governed by the equation of motion

\[
0 = \frac{d\phi(R, \tau)}{d\tau} = \frac{\partial \phi(R, \tau)}{\partial \tau} + v(R_i) \cdot \nabla \phi(R, \tau).
\] (17)

The velocity field in this equation is the same as in Eq. (14), namely

\[
v(R_i) = n(R_i) f_\Delta(R_i) = \frac{\nabla \phi(R, \tau)}{\left| \nabla \phi(R, \tau) \right|} f_\Delta(R_i).
\] (18)

The second equality is obtained by using the property (ii) of the list above.

In the following section we apply this algorithm, based on the LSM, to compute the MPDs in a liquid system.

IV. NUMERICAL IMPLEMENTATION

A. Choice of the initial volumes

To apply the LSM we need to choose a starting set. In the two cases of pure water or sodium diluted in water we locate the initial domains \( \Delta_i \) in correspondence to the local maxima of the oxygen-oxygen SDF or of the sodium-oxygen RDF, respectively. This difference is related to the fact that the SDF of a sodium ion is spherically symmetric.

The two-particle SDF \[29\], employed here and calculated around a central (0th) water molecule, is proportional to the probability of finding an oxygen atom in \( R \) irrespective of the position of all other oxygen atoms \[1\], given a certain configuration \( \{R_0, R_H^1, R_H^2\} \) of the central molecule:

\[
\rho^{(2)}(R) = \rho^{(2)}(R | R_0, R_H^1, R_H^2) = N(N - 1) \int dR^{N-1} \rho \left( R, R^{N-1} | R_0, R_H^1, R_H^2 \right),
\] (19)

where the probability density under the integral sign is

\[
\rho \left( R, R^{N-1} | R_0, R_H^1, R_H^2 \right) = \rho \left( R, R^{N-1} \right)
\]

\[
= \int dR_H^{1,N} dR_H^{2,N} \frac{\rho^{can} \left( R_0, R, R^{N-1}, R_H^1, R_H^2, R_H^{1,N}, R_H^{2,N} \right)}{P_m \left( R_0, R_H^1, R_H^2 \right)}
\] (20)

exactly as in Eq. (10), with \( P_m \left( R_0, R_H^1, R_H^2 \right) \) given by Eq. (11). Note that \( \rho^{(2)}(R) \) is now a function of the three cartesian coordinates \( R \) and can therefore be visualised in 3D space, as we will show below.
By employing Bader analysis [30], we locate maxima, and domains around them, of the two-particle SDF in Eq. (19). In doing that, we apply an analysis method designed for the study of electronic probability densities to the study of a liquid. Bader analysis partitions the space in regions assigned to the maxima of the density. From these regions starts the search of the MPDs which is not a partition of the space and can result in overlap between domains or partial filling of the spaces. In actual calculations, to fully identify the initial Bader domains, given an initial $\mathbf{R}'$, we compute the vector $\mathbf{R} = \mathbf{R}' - \mathbf{R}_0$, where $\mathbf{R}_0$ is the origin of our reference system, and we evaluate the gradient of $\rho^{(2)}(\mathbf{R})$ at that point. Then we construct a steepest-ascent path in 3D space whose tangent in every point is the direction of the gradient. When the maximum is reached, all points of that path are associated to a Bader volume. Another point $\mathbf{R}'$ is then chosen and the procedure is iterated until all space points have been assigned to a specific maximum. Indeed, starting from different points, the same maximum can be reached thus the Bader volume has to be updated.

This approach fails in the case of the spherically symmetric sodium-oxygen SDF. Therefore, in this case we proceed differently. We take as initial sets few spheres of different radii, generally smaller than 2.5 Å, whose centres are approximately located at a distance from sodium corresponding to the two main maxima of the sodium-oxygen RDF, i.e. at $\sim 2.5$ and $\sim 4.5$ Å from the sodium ion (see for instance Fig. 17).

### B. Algorithm

According to the steps described in the previous section, we

1. define the initial domain $\Delta_0$;
2. construct a level set function. Our choice [28] is

$$
\phi(\mathbf{R}, 0) = \begin{cases} 
-1 + \exp[-d(\mathbf{R})/\sigma] & \text{if } \mathbf{R} \in \Delta \\
0 & \text{if } \mathbf{R} \in \partial \Delta \\
1 - \exp[-d(\mathbf{R})/\sigma] & \text{if } \mathbf{R} \not\in \Delta
\end{cases}
$$

(21)

where $d(\mathbf{R})$ is the shortest distance of the point $\mathbf{R}$ from $\partial \Delta$ and $\sigma$ is a parameter to be chosen;
3. calculate $P^{(1)}(\Delta)$ from Eq. (7);
4. calculate $f_\Delta(\mathbf{R})$ from Eq. (15);
5. evolve the level set function according to Eq. (17) (its numerical “equivalent” used in the implementation is given in Eq. (A27));
6. go back to point 3. and iterate until the variation of the probability $P^{(1)}(\Delta)$ is smaller than a certain threshold $\delta$, i.e. $|P^{(1)}(\Delta_t) - P^{(1)}(\Delta)| \leq \delta$.

The fictitious time step $d\tau$ to be used in our calculations has resulted to be 10.0, with a value for $\sigma$ of 0.1 Å while the threshold to monitor the convergence of the algorithm has been chosen to be $\delta = 10^{-7}$. A good grid spacing for the 3D visualisation of the MPDs has resulted to be 0.2 Å.

### V. CASE STUDIES

In the first solvation shell of pure water at suitable conditions defined later on, Bader domains are shown in Fig. 1 along with $\rho^{(2)}(\mathbf{R})$. If we restrict ourselves to the four highest maxima of $\rho^{(2)}(\mathbf{R})$ computed for water, Bader analysis identifies the domains $\{\Delta_i\}_{i=1,4}$ that are represented by the volumes enclosed by the grids in Fig. 1. The filled areas are shown for reference and represent the function $\rho^{(2)}(\mathbf{R})$. For each domain, the probability of finding one, and only one, particle $P^{(1)}(\Delta_i)$ is calculated from Eq. (7) and then maximised, by following the procedure illustrated above. The final domain is the MPD available to each particle surrounding the central water molecule.

The applications of the LSM are illustrated below. Before doing that, we find useful to stress two points. (A) The results of our analysis do not depend on the initial set chosen. To that end, let us look at the result of the optimisation performed using two different initial conditions for maxima of equivalent physical meaning. The first type of initial condition is a Bader domain (green areas in Fig. 2 left). The second, for different but equivalent maxima, is a sphere, approximately centred at the positions of the maxima of the SDF (blue areas in Fig. 2 left). Fig. 2 (right) shows that the shape of the MPDs is independent of the initial conditions.
FIG. 1. The filled areas represent the two-particle SDF (rendering of contour surface 0.17 of $\rho^{(2)}(R)/\rho_0$). The grids enclose the Bader domains, namely the volumes which are associated to the four highest maxima of the two-particle SDF obtained by the application of Bader analysis.

(B) When Bader analysis is applied in order to identify the initial domains, the space around the central particle (water molecule in our case) is partitioned as far as the two-particle SDF shows well defined local maxima. When the liquid approaches a random distribution Bader analysis becomes inefficient. In the case of water at density $\rho_0 = 1.0 \text{ g/cm}^3$ and $\rho = 1.23 \text{ g/cm}^3$ presented in two following sections, we are able to identify and analyse domains up to within the third solvation shell ($6 - 8 \text{ Å}$).

A. Liquid water at $\rho_0 = 1.0 \text{ g/cm}^3$

The MD trajectory of 150 ps for liquid water at room temperature is generated by employing the TIP4P model. The system is composed of 4096 molecules in a cubic box with side length 49.7 Å. Periodic boundary conditions are used throughout.

The 3D two-particle SDF has been calculated from the MD trajectory and, from the application of Bader analysis, the space around a central water molecule is partitioned in 15 regions within a distance of about 8 Å from the central oxygen atom. The resulting 15 MPDs give information on the structure of water up to within the third solvation shell, as we will now show.

Bader domains and MPDs in the first solvation shell around the central water molecule are shown in Fig. 3. These domains clearly identify the regions of space associated to the four water molecules hydrogen-bonded (H-bonded) to the central molecule, arranged according to the typical tetrahedral structure of the first solvation shell of water. Domains 1 and 2 contain the water molecules H-bonded to the hydrogen atoms, domains 3 and 4 contain those molecules which are H-bonded to the oxygen atom. It is worth noting that, as already discussed, the domains do not partition the space after optimisation because they are allowed to overlap, as is the case of the domains 3 and 4 in Fig. 3 or to be separated by empty spaces. In particular, we will show that within a radius of about 6 Å from the central molecule relatively large cavities separate MPDs from one another. These cavities are mainly the result of the repulsive interactions and of the internal structure of the water molecules. Even if we are looking at the MPDs enclosing the oxygen atoms, the probability distribution that we analyse indeed accounts for the full molecular structure of water.

In general we can classify the MPDs around the central water molecule according to the value of the probability $P^{(1)}(\Delta)$ in each domain. The domains labeled 1 and 2 are associated to a probability of about 0.9 of finding one and only one particle enclosed in those regions; this probability is smaller in the domains labeled 3 and 4, namely 0.78. This difference can be explained as follows. Domains 1 and 2 enclose the oxygen atoms of the two water molecules
FIG. 2. Left: different initial conditions for two sets of equivalent domains around a water molecule. Two domains are located in the region of the first solvation shell corresponding to the two oxygen atoms accepting the hydrogen bond; two domains appear in the region in the first solvation shell occupied by the two water molecules donating the hydrogen bonds. The green areas are Bader domains, whereas the blue areas are spheres with centres located at the positions of the maxima of the two-particle SDF. Right: the domains located at equivalent positions are identical after optimisation.

FIG. 3. MPDs (represented by the filled areas) labeled 1, 2 (blue areas) and 3, 4 (red areas) around the central water molecule. The initial Bader domains are represented as grids.

H-bonded to the hydrogen atoms of the central water molecule, namely HO⋯H\textsubscript{central}. Domains 3 and 4 enclose the oxygen atoms of the molecules that are H-bonded to the central one as OH⋯O\textsubscript{central}. Therefore, in the latter case, the oxygen atoms are not directly bonded to the central molecule, thus resulting more mobile.

If the occupancy probability of domains 1 to 4 is compared to the ideal gas, these probabilities are much higher than the optimised value \( P^{(1)}(\Delta^*) = 0.37 \) in absence of interactions (see Appendix B). Also, it is reasonable to expect...
that such probability decreases as the distance from the central water molecule increases due to the weakening of the bonding.

Since the domains 1 to 4 enclose the oxygen atoms of the water molecules that are H-bonded to the central molecule, we can define an estimate of the number \( n_{\text{HB}} \) of hydrogen bonds (HBs) formed by a water molecule. Indeed the domains indicate only the positions of the oxygen atoms, and therefore, domains 3 and 4 cannot, strictly speaking, be used to estimate the presence of HBs. This is because in those regions the hydrogen atoms (not the oxygens) form HBs with the central water molecule. However, the H-bonded hydrogens are close to the oxygens occupying the domains 3 and 4. If we take a high one-particle occupancy probability as an indication of the presence of a HB, then its value can qualitatively measure the fraction of HBs present. Taking that the four closest domains do not superpose (a hypothesis almost always satisfied, but certainly an approximation \([32]\)), we can sum up their probabilities to have an indication of how many HBs are formed around a water molecule:

\[
 n_{\text{HB}} = \sum_{i=1}^{4} P^{(1)}(\Delta_i). \tag{22}
\]

The usual distance criterion for determining the existence of a HB between two water molecules is automatically accounted for by restricting the number of domains in the sum to the four that are the closest to the central molecule. In a transient configuration, when one hydrogen bond is broken before another is formed, it is plausible to assume that the domain is occupied by a number of water molecules different from one. According to Eq. \((22)\), we find that the average number of HBs formed by a water molecule is \( n_{\text{HB}} = 3.37 \), in agreement \([32, 33]\) with the values reported in the literature of \( 3.1 - 3.6 \).

The domains 5 to 8 are shown in Fig. 5. They are arranged on both sides of the plane defined by the central molecule in front of the hydrogen atoms and are symmetric with respect to this plane. The probability of finding only one water molecule inside each domain is reduced, compared to the previous case, to 0.63. The same value is calculated after the optimisation of the domains 9 - 10, shown in Fig. 6 and 11 - 12, shown in Fig. 7. These sets of domains have similar shape after the optimisation and the volume is significantly reduced if compared to the corresponding Bader domains.

In the context of the MPDs, solvation shells can be defined by calculating the oxygen-oxygen RDF resolved in each optimised domain. In Fig. 8 we show the RDF computed within a MPD (lower panel) and compared to the total RDF (upper panel). The \( x \)-axis has been divided in three regions, each representing a solvation shell around the central water molecule. In particular, the first shell is identified as the region where the RDF in domains 1 to 4 is significantly non-zero. The extent of this region is up to about 3.5 Å. The MPDs labeled from 5 to 12 (corresponding to the blue and magenta lines in Fig. 8) only partially occupy the cavities between the domains 1 to 4 of the first solvation shell, but they are mainly found in the second solvation shell.

The coordination number in the first solvation shell is 5.1, if computed as the integral of the total RDF up to 3.5 Å, and 4.5, if computed as the sum of the integrals of the partial RDFs. An extra-particle \([34]\) appears in the first solvation shell, which does not occupy one of the four domains of the tetrahedron, but it is delocalised in the interstitial domains 5 to 12. Particularly interesting is the distribution of particles in domains 5 to 12 since the partial RDFs, although very flat, show peaks at about 3.5 Å. Similar maxima have been already observed by Svishchev and...
FIG. 5. MPDs (represented by the filled areas) labeled 5, 6 (cyan areas) and 7, 8 (orange areas) around the central water molecule. The initial Bader domains are represented as grids.

FIG. 6. MPDs (represented by the filled areas) labeled 9, 10 (violet areas) around the central water molecule. The initial Bader domains are represented as grids.

Kusalik [2] from calculation of RDFs resolved in angle. This feature of the RDFs suggests the presence of water molecules in non-tetrahedral directions but still within the first solvation shell. The coordination numbers in the first solvation shell estimated from the integral of the total RDF and of the partial (within each MPD) RDFs slightly differ. In particular, the value of 4.5 from our analysis is smaller than the standard estimate of 5.1. This difference occurs as consequence of requesting the exclusive occupancy of the MPDs by only one oxygen atom. The physical picture obtained by the analysis of the MPDs is somehow different from the picture in terms of reduced probabilities and thus can be used to complement more standard analysis techniques.

The MPD labeled 13 in Fig. 9 is located in the second solvation shell as well. From Fig. we can identify the limit of such region with the zero of the oxygen-oxygen RDF associated to this domain (green curve in the figure). Also in this case the volume has been significantly reduced by the optimisation procedure, however the probability of finding
FIG. 7. MPDs (represented by the filled areas) labeled 11, 12 (ochre areas) around the central water molecule. The initial Bader domains (completely symmetric with respect to the plane of the molecule and to its perpendicular plane) are represented as grids.

| Domain label | \( \langle N \rangle_{\text{before}} \) | \( \langle N \rangle_{\text{after}} \) | Volume (Å\(^3\)) | Distance (Å) | \( P^{(1)}(\Delta^*) \) |
|--------------|-----------------|-----------------|------------------|-----------|--------------|
| 1, 2         | 1.2             | 1.1             | 14.9             | 2.8       | 0.9, 0.91    |
| 3, 4         | 1.6             | 1.0             | 15.2             | 2.6       | 0.78         |
| 5 - 8        | 2.0             | 1.2             | 15.8             | 4.2       | 0.63         |
| 9, 10        | 1.8             | 1.1             | 15.9             | 4.3       | 0.63         |
| 11, 12       | 5.9             | 1.1             | 15.8             | 4.4       | 0.63         |
| 13           | 3.0             | 1.8\(^{(a)}\)   | 17.2             | 4.5       | 0.37\(^{(a)}\) |
| 14, 15       | 4.0             | 0.6\(^{(a)}\)   | 21.1             | 6.5       | 0.61\(^{(a)}\) |

TABLE I. List of the MPDs (first column), average number of particles in each volume before (Bader analysis, second column) or after the optimisation (third column), volumes of the MPDs (fourth column), distances of the centres of mass of the MPDs from the central oxygen atom (fifth column) and the one-particle occupancy probability (sixth column). In the same line of the first column we have put equivalent domains. \(^{(a)}\) The values of \( \langle N \rangle_{\text{after}} \) and \( P^{(1)}(\Delta^*) \) for the volumes 13 and 14/15 are in italic to indicate that the optimisation procedure seems to be in these cases less efficient than for the other domains: in the first case, the optimisation appears not able to split the domains in two parts; in the second case, the density of the liquid is probably too flat, thus preventing from an efficient analysis.

only one particle enclosed by this volume remains low, namely around 0.37.

In order to understand this result, we calculate the average number of particles that are found in the MPDs. These results are listed in table I. In all MPDs discussed so far, from label 1 to label 12, after the optimisation the average number of particles \( \langle N \rangle_{\text{after}} \) enclosed by the optimised volume is very close to one. In the case of domain 13, this number is 1.8. It is likely that the proposed implementation of the LSM is not able to split the domain in two parts or that the probability distribution is very flat thus preventing from an efficient optimisation.

Table I shows the volumes associated to each MPDs and their distance from the oxygen atom of the central water molecule, expressed in terms of the distance of their centres of mass. In particular, we notice the clear identification of three groups of domains, whose distances from the central molecule are below 3 Å, between 4 and 5 Å, above 6 Å. We also observe the increase of the volume of the domain itself as its distance from the central oxygen increases. The correlation to the central water molecule is reduced, then the surrounding molecules are more mobile and occupy
larger volumes.

The values of the one-particle occupancy probability for the MPDs and their volumes from Table I can be compared with the corresponding results calculated for the ideal gas in Appendix B. We observe that the probability in the absence of interactions, as also stated above, is 0.37. This value is much smaller than those calculated for water surrounding a water molecule, ranging from about 0.6 to more than 0.9. The interactions among the particles of the liquid indeed stabilise their distribution and it is in particular the repulsion that prevents the occupation of a MPD by more than one particle (that would lead to the decrease of the value of the probability). Moreover, the average volumes per particle $v_0$ calculated for water at $\rho_0$ is 30.0 Å$^3$, which is also the volume of a MPD in the ideal gas at the same density $\rho_0$. The volume predicted for the ideal gas is larger than all those calculated for water, listed in Table I, suggesting that interactions make the particles less mobile than what is expected in a non-interacting situation. These conclusions are very general, as we will see that they apply also in the following cases of water at higher density and of the sodium ion in water.

The third solvation shell is partially occupied by the MPDs labeled 14 and 15, shown in Fig. 9. Also in this case the final volumes appear significantly reduced if compared to the initial Bader domains. In this case, the probability of finding only one particle within the domains largely increases during the optimisation process. This is clearly shown in Fig. 10 by observing the histogram associated to the probability, before and after the optimisation, of finding only one particle inside each domain. The initial value of $P^{(1)}(\Delta)$ for domains 14 and 15 is very low, around 0.02, but it reaches the value 0.61 after the optimisation. However, as shown in Table II the average number of particles in the MPDs 14 and 15 is 0.6. This result can be interpreted by referring to the histograms in Fig. 10. In all cases the optimisation has the effect of increasing the probability of occupation of the domain by only one particle, while reducing the probability of occupation by zero or two particles. This does not happen in the case of domains 14 and 15, as also the probabilities of zero and two particles increase.

A similar analysis will be presented in the following sections to investigate the distribution of water molecules around a central water molecule in liquid water at density 1.23 g/cm$^3$ and around a solvated Na$^+$ ion. In particular, the results on the analysis of the structure of water at higher density will be presented in comparison with the results of this section.
FIG. 9. MPDs (represented by the filled areas) labeled 13 (light purple area) and 14, 15 (turquoise areas) around the central water molecule. The initial Bader domains are represented as grids.

FIG. 10. Probabilities of finding only \( n = 0, 1, 2 \) in each domain, before (black) and after (red) the optimisation. Only the results for one domain of each kind are shown, but the probabilities are almost identical for domains of the same kind.

B. Liquid water at \( \rho = 1.23 \text{ g/cm}^3 \)

The MD trajectory of 150 ps for liquid water at room temperature is generated by employing the TIP4P [31] model. The system is composed of 4096 molecules in a cubic box with side length 46.3 \( \text{Å} \).

Bader analysis is applied to the two-particle SDF computed for liquid water at the density \( \rho = 1.23 \text{ g/cm}^3 \) and 11 domains are identified, in contrast to the previous case where 15 domains were defined by Bader analysis. The shapes and positions of the initial domains are shown in Fig. [11] whereas the MPDs are shown in Fig. [14]. Before commenting on the MPDs, it is worth discussing the initial configurations. As expected, the first solvation shell is unaffected [10, 35] by the increase of density, namely domains 1, 2 and 3, 4 are similar to the previous situation. The
remaining domains are arranged closer to the central molecule and they are more localised in space.

The main difference observed in Fig. 11 is that the domains labeled from 5 to 10 in Figs. 5 and 6 merged in two domains, as shown in Fig. 12 where we compare the two results of Bader spatial decomposition. This comparison explains why the total number of domains is reduced from 15, at low density, to 11, at higher density. This is indeed an expected results, since the water molecules are more “packed” at higher density and the maxima of the $\rho^{(2)}(\mathbf{R})$ are less sharp. These two domains will be labeled 5 and 6 and shown as cyan areas in Figs. 11 and 14. Moreover, the size of the domains labeled 11 to 13 (labeled in this case 7, 8, ochre areas, and 9, light purple area, in Figs. 11 and 14) is reduced already from the initial configuration.

The main difference observed in Fig. 11 is that the domains labeled from 5 to 10 in Figs. 5 and 6 merged in two domains, as shown in Fig. 12 where we compare the two results of Bader spatial decomposition. This comparison explains why the total number of domains is reduced from 15, at low density, to 11, at higher density. This is indeed an expected results, since the water molecules are more “packed” at higher density and the maxima of the $\rho^{(2)}(\mathbf{R})$ are less sharp. These two domains will be labeled 5 and 6 and shown as cyan areas in Figs. 11 and 14. Moreover, the size of the domains labeled 11 to 13 (labeled in this case 7, 8, ochre areas, and 9, light purple area, in Figs. 11 and 14) is reduced already from the initial configuration.

Also in this case the MPDs can be grouped according to the corresponding final probability of enclosing only one particle. The domains 1, 2 are associated to a high probability of 0.93 and the domains 3, 4 to a probability of 0.78. These values are higher than in the case of lower density, resulting in an increase of the calculated average number of HBs per molecule. If we use again Eq. (22) to estimate this value, we find $n_{HB} = 3.41$. We observe a qualitative difference between the MPDs 3 and 4 determined at density $\rho_0$, shown in Fig. 4 and the equivalent domains at higher density $\rho = 1.23 \rho_0$, shown in Fig. 13. In this case, we observe that the domains do not overlap, thus leading to a stronger localisation of the molecules enclosed in these regions. Apart from these differences, the optimised domains in the first solvation shell are not very much affected by the change of density. This observation agrees with the literature, namely the MPDs analysis confirms that the first solvation shell is quite rigid under density increase.

The probability of finding only one particle in the MPDs 5, 6 (cyan areas in Fig. 14) is 0.74, in the MPDs 7, 8 (ochre areas in Fig. 14) is 0.72, in the MPD 9 (light purple area in Fig. 14) is 0.68 and in the MPDs 10, 11 (turquoise areas in Fig. 14) is 0.70. The optimisation procedure seems to be more effective in this case of higher density, as the average number of particles found in each MPD is very close to one for all domains, as shown in table II. This effect can be interpreted as a stabilisation of the molecules in the second solvation shell due to the packing imposed by the larger density. The volume per particle in this case is 24.2 Å$^3$, that is also the volume of a MPD in a system at the same density $\rho = 1.23$ g/cm$^3$ but without interactions. We see again, as discussed in the previous section, that when interactions are taken into account, the dimensions of the MPDs is largely reduced.

Very interesting is the behaviour of the MPDs 5, 6, which are found at a distance of 3.7 Å from the central oxygen atom, as shown in table II. In the previous analysis the MPDs were arranged in three groups, namely with distances below 3 Å, between 4 and 5 Å, above 6 Å from the central oxygen. In this case we observe that the first group does not change and the distance of the MPDs 1 to 4 remains below 3 Å. The second group of MPDs presents now different features, since the domains labeled 5, 6 are found at a distance of 3.7 Å and the domains labeled 7 to 9 at a distance between 4 and 5 Å, exactly as the second group of domains at density $\rho_0$. The last group of domains is found at a
distance slightly smaller than 6 Å.

In order to further analyse the main difference between the two cases, namely the appearance of the domains at a distance of 3.7 Å from the central water molecule, we show in Fig. 15 the partial RDFs calculated inside domains 5, 6. A pronounced peak (blue line in the figure) appears at around 3.3 Å, contributing to the shoulder in the total oxygen-oxygen RDF at the same distance (highlighted in the figure by the circle). In the previous analysis we showed that all domains in these regions, namely those labeled 5 to 12, have similar partial RDFs (blue and magenta lines in Fig. 8). Instead we observe here that the partial RDFs calculated inside the domains 5, 6 show a more pronounced peak than the domains 7, 8. In general, we can observe that the MPDs 5 to 8 enclose the four molecules [10] that in non-tetrahedral directions [2, 37] mainly contribute to modifications of the second solvation shell at increasing density. The hypothesis [10] that, as the density increases, the second solvation shell does not continuously collapse on the first shell but interstitial molecules, that are not H-bonded to the first shell molecules, get closer to the central water is consistent with our results.

The MPD 9 shows some differences with respect to the corresponding domain 13 at ρ₀. In particular, its initial
FIG. 14. MPDs for water at \( \rho = 1.23 \text{ g/cm}^3 \). As in Fig. 11 only one domain of those labeled 5 and 6 is shown (cyan region), in order to make the central water molecule visible.

| Domain label | \( \langle N \rangle_{\text{before}} \) | \( \langle N \rangle_{\text{after}} \) | Volume (\( \AA^3 \)) | Distance (\( \AA \)) | \( P^{(1)}(\Delta^*) \) |
|--------------|----------------|----------------|------------------|----------------|------------------|
| 1, 2         | 0.7            | 0.9            | 12.6             | 2.7            | 0.93             |
| 3, 4         | 1.7            | 1.0            | 12.0             | 2.8            | 0.78             |
| 5, 6         | 5.1            | 1.3            | 14.5             | 3.7            | 0.74             |
| 7, 8         | 2.7            | 1.3            | 15.0             | 4.0            | 0.72             |
| 9            | 2.4            | 1.0            | 15.8             | 4.7            | 0.68             |
| 10, 11       | 5.0            | 0.9            | 16.1             | 5.6            | 0.70             |

TABLE II. List of the MPDs (first column), average number of particles in each volume before (Bader analysis, second column) or after the optimisation (third column), their volumes (fourth column), distances of their centres of mass from the central oxygen atom (fifth column) and the one-particle occupancy probability (sixth column). In the same line of the first column we have put equivalent domains.

...shape is more localised in space. The LSM produces an optimised MPD that on average is occupied by exactly 1 particle (see table II).

The optimised domains 10, 11 are qualitatively similar to the corresponding domains 14, 15 in Fig. 9. However, a more detailed analysis shows that the average occupation is closer to one than in the previous case and they are closer to the central molecules, as shown in table II.

C. \( \text{Na}^+ \) in water

The MD trajectory of 110 ps for a sodium ion in liquid water, at room temperature, is generated by employing the TIP4P model for water. The system is composed of 1024 molecules in a cubic box with side length 31.5 \( \AA \).

In the case of a sodium ion in liquid water, MPDs are determined as those regions of space where the probability of finding one, and only one, oxygen atom (and, consequently, one water molecule, irrespective of its orientation) in the neighbourhood of the central \( \text{Na}^+ \) is maximum. The two-particle SDF calculated for \( \text{Na}^+ \) surrounded by water molecules is spherically symmetric and it does not contain more information than the ion-oxygen RDF. Therefore, MPDs will have the same spherical distribution around the central ion. The position and shape of the initial domains can be chosen totally arbitrarily, since Bader analysis is not efficient in this situation, where the local maxima of the two-particle SDF cannot be properly located (maxima of the density are distributed on a sphere, they are not isolated points in 3D space). Also, if the initial domains are chosen within a distance of less than about 2.2 \( \AA \), where the RDF (see Fig. 17) is very small, the optimisation procedure is not efficient and the MPDs cannot be identified. This problem
TABLE III. List of properties of the MPDs calculated for \(\text{Na}^+\) in water. In the first column we list the position of the domain, in one of the two solvation shells around the sodium ion, in the second we calculate the volumes occupied by the domain, in the third column we report the distance of the centre of mass of the domain from the central ion, in the fourth column we show the solid angle occupied by the domain and in the fifth column, the coordination number associated to it. In parenthesis, we compare the value of the coordination number determined by integrating the RDF up to \(R = 3.3, 5.6\) Å. In the sixth column the values of the one-particle occupancy probability are given.

| Shell | Volume (Å\(^3\)) | Distance (Å) | Ω (deg\(^2\)) | \(n_Ω\) | \(P^{(1)}(\Delta^*)\) |
|-------|------------------|--------------|----------------|--------|------------------|
| I     | 12.0             | 2.3          | 135.6          | 5.3 (5.9) | 0.78             |
| II    | 16.2             | 4.4          | 39.9           | 18.1 (19.3) | 0.62             |

is related to the fact that in empty regions, where the probability of finding one particle vanishes, a small variation of the region itself does not change this probability, thus fulfilling the optimisation condition \(|P^{(1)}(\Delta) - P^{(1)}(\Delta)| \leq \delta\) without an effective modification of the domain. Here, we have considered only domains \(\Delta\) that are enclosed in the region of non-zero probability density. We then choose, as initial domains, spheres of different radii, randomly located around the central ion and at distances between 2.5 Å and 4.5 Å from it. The optimisation procedure is applied to these initial conditions and two sets of optimised domains are identified, associated to the first and second solvation shells. The corresponding MPDs are shown in Fig. 16 as blue and red grid volumes, with distances from the central ion of 2.3 and 4.4 Å, respectively. These values are listed in table III along with the other properties associated to the two MPDs. The maximum probabilities of finding only one particle inside each region are 0.78 and 0.62 (in analogy to the previous sections, the lower value corresponds to the MPDs further from the central ion).

Fig. 16 shows that the shape of the inner domain, which is associated to the first peak of the RDF, is more localised in space, as it is expected in connection to the fact that the first peak of the RDF is narrower than the second one. Consequently, the second domain spreads further in space from the central ion.

Table III shows the volumes occupied by each domain. The two values are much smaller than the value for the corresponding ideal gas, which is 30.5 Å\(^3\), as also happens in the cases of water. The coordination number in the two shells is determined as the ratio of the solid angle occupied by the domain and the total solid angle. The solid angle \(Ω\) and the coordination number \(n_Ω\) thus determined are shown in table III. In particular, the coordination numbers \(n_Ω\) are in very good agreement with those determined by integrating the RDF up to the distances 3.3 and 5.6 Å, for the first and second solvation shell, respectively. As previously discussed in the case of water, slightly smaller values are determined by our analysis, if compared to the integral of there RDF. Furthermore, we can reconstruct the total RDF, by determining its value inside the MPDs. The comparison between the total RDF and the partial ones (in
FIG. 16. MPDs in the first and second solvation shells of Na\(^+\). From a snapshot of the trajectory, some water molecules in the vicinity of the central ion are also shown. We see that each domain is occupied by only one molecule (indicated by the arrows).

FIG. 17. Total and partial (calculation restricted to each given domain) Na\(^+\)-oxygen RDF. The blue and red curves are determined by calculating the RDF in the domains shown in Fig. 16 and by multiplying it by the corresponding \(n_\Omega\) (from table III).

VI. CONCLUSIONS

We extended the study of the structure of a liquid system by the identification of MPDs, borrowing the idea from what is now routinely done in electronic structure analysis, and we developed a first application to few case studies. We derived the formalism necessary to define the probability of finding one and only one particle in a certain region of space \(\Delta\) and we developed an algorithm for the optimisation of this region under the request that the probability is maximised. The geometric optimisation is formulated in terms of shape derivatives, thus allowing to use the LSM to solve the optimisation problem.
The case study obtained the features of the MPDs for water at different densities and for a solvated sodium ion in water. A qualitative analysis of the domains in water has allowed us to describe the modification of the second solvation shell by increasing pressure. More quantitative observations have also illustrated the presence of water molecules in the interstitial spaces of the first solvation shell, that are not H-bonded to the central molecule. These molecules contribute to the modification of the second solvation shell of water at increasing density. In the case of a sodium ion in water, we reconstructed the oxygen-ion RDF from the partial contributions evaluated inside the optimised domains and we determined the coordination numbers associated to these domains. We used a criterion to identify the solvation shells, based on the identification of regions where the partial RDFs, calculated within the MPDs, are close to zero, rather than on the identification of the minima of the total RDF.

We investigated the properties of the MPDs that are determined from the maximisation of the one-particle occupancy probability. Along similar lines one could analyse probabilities associated to more than one particle or focus on other atoms (hydrogens, for instance). This could be used to give a many-body definition of solvation shells.

The proposed method results to be an efficient tool to complement the standard analysis techniques used in the study of the structure of a liquid system and to give a more detailed 3D resolution of the organisation of the space around a given molecule, or atom or ion. As it clearly appears from the results presented throughout the paper, we are able to observe properties that are in agreement with other kind of analysis. Yet, since we look at the problem from a new perspective, we see the possibility of interesting developments.

ACKNOWLEDGEMENTS

F. A. and R. V. acknowledge the support received from the European Science Foundation (ESF) for the activity entitled “Molecular Simulations in Biosystems and Material Science”. G. C. is grateful to E. K. U. Gross for his hospitality at the Max-Planck-Institute in Halle and his support all along the long gestation of this work. The authors thank D. Borgis for fruitful discussions at the initial stages of this work.

Appendix A: The level set method

We present here in detail the LSM, defining the shape derivative of $P^{(1)}(\Delta)$ and introducing the level set function.

1. Shape derivative

$P^{(1)}(\Delta)$ belongs to a specific class of what is normally called a set function. We now introduce few analysis tools used when treating in general set functions.

The shape derivative of a set function of the form

$$F(\Delta) = \int_{\Delta} d\mathbf{R} f(\mathbf{R})$$

is defined as the variation with respect to the integration domain $\Delta$. In other words $F(\Delta)$ can be written in terms of a “density”, $f(\mathbf{R})$, and, in the following section, we give the explicit definition of the $f(\mathbf{R})$, interesting us, from Eq. [1].

If the deformation law of $\Delta$ is given as

$$\Delta \rightarrow \Delta_\tau = \{ \mathbf{R}_\tau = \mathbf{R} + \mathbf{v}(\mathbf{R})d\tau | \mathbf{R} \in \Delta \}$$

where $d\tau$ is a fictitious time step and $\mathbf{v}(\mathbf{R})$ is a velocity field, the time derivative of $F(\Delta)$ can be calculated according to

$$\frac{dF(\Delta)}{d\tau} = \lim_{d\tau \rightarrow 0} \frac{1}{d\tau} \left[ \int_{\Delta_\tau} d\mathbf{R}_\tau f(\mathbf{R}_\tau) - \int_{\Delta} d\mathbf{R} f(\mathbf{R}) \right].$$

Performing, in the first term in square brackets, the change of variable

$$\mathbf{R}_\tau = \mathbf{R} + \mathbf{v}(\mathbf{R})d\tau, \quad d\mathbf{R}_\tau = d\mathbf{R} \left( 1 + \nabla \cdot \mathbf{v}(\mathbf{R})d\tau + \mathcal{O}(d\tau^2) \right)$$

and expanding, for small time increment $d\tau$, $f(\mathbf{R}_\tau)$ around $\mathbf{R}$

$$f(\mathbf{R}_\tau) = f(\mathbf{R}) + \mathbf{v}(\mathbf{R}) \cdot \nabla f(\mathbf{R})d\tau,$$
Eq. (A3) becomes
\[
\frac{dF(\Delta)}{d\tau} = \int_{\Delta} d\mathbf{R} \nabla \cdot (\mathbf{v}(\mathbf{R}) f(\mathbf{R}))
\]
(A6)
\[
= \int_{\partial\Delta} ds \mathbf{n}(\mathbf{R}) \cdot \mathbf{v}(\mathbf{R}) f(\mathbf{R}) \equiv \mathcal{D}_S F(\Delta)
\]
(A7)
where the divergence theorem has been used to derive Eq. (A7) from Eq. (A6). In the last line \(ds\) is the surface element of the boundary \(\partial\Delta\) of \(\Delta\) and \(\mathbf{n}(\mathbf{R})\) is the unit vector normal to the surface at \(\mathbf{R}\). Eq. (A7) defines the shape derivative, indicated by the symbol \(\mathcal{D}_S\), of \(F(\Delta)\) and expresses the variation of \(F(\Delta)\) with respect to the variation of \(\Delta\). The change in the domain \(\Delta\) is expressed in terms of a global deformation of the boundary \(\partial\Delta\), which, in turn, is determined by the velocity field \(\mathbf{v}(\mathbf{R})\). For our purpose, that is the maximisation of \(F(\Delta)\), we can choose the field \(\mathbf{v}(\mathbf{R})\) such that \(\mathcal{D}_S F(\Delta) \geq 0\) and follow the evolution of such a field up to find the final \(\mathcal{D}_S F(\Delta) = 0\). The condition \(\mathcal{D}_S F(\Delta) \geq 0\) is automatically satisfied if we choose the velocity field as
\[
\mathbf{v}(\mathbf{R}) = \mathbf{n}(\mathbf{R}) f(\mathbf{R}).
\]
(A8)

2. Shape derivative for the MPDs

The comparison between Eqs. (1) and (A1) shows that when we express the probability \(P^{(1)}(\Delta)\) in terms of a density function similar to \(f(\mathbf{R})\), such density depends on \(\Delta_c\), i.e. on \(\Delta\). Therefore, when calculating the variations of \(F(\Delta)\) from Eq. (A1) with respect to \(\Delta\), we need to include also variations of \(f(\mathbf{R})\) with respect to \(\Delta_c\), i.e. \(\Delta\). In our case, then, the transformation in Eq. (A4), restricting ourselves to the case \(\mathbf{v}_j(\mathbf{R}) = \mathbf{v}(\mathbf{R}_j)\), becomes
\[
\mathbf{R}_{j,\tau} = \mathbf{R}_j + \mathbf{v}(\mathbf{R}_j) d\tau \quad \text{with } j = 1, \ldots, N
\]
(A9)
and the volume elements are
\[
d\mathbf{R}_{j,\tau} = d\mathbf{R}_j \left(1 + \nabla \mathbf{R}_j \cdot \mathbf{v}(\mathbf{R}_j) d\tau\right).
\]
(A10)
The time derivative of the probability \(P^{(1)}(\Delta)\) under the variation \(\Delta \to \Delta_c\) is
\[
\frac{dP^{(1)}(\Delta)}{d\tau} = \lim_{d\tau \to 0} \frac{1}{d\tau} \sum_{i=1}^{N} \left[ \int_{\Delta_c,\tau} d\mathbf{R}_i, \tau \left( \prod_{\ell \neq i} \int_{\Delta_c,\tau} d\mathbf{R}_\ell \right) \rho(\mathbf{R}_i^N) - \int_{\Delta} d\mathbf{R}_i \left( \prod_{\ell \neq i} \int_{\Delta_c} d\mathbf{R}_\ell \right) \rho(\mathbf{R}_i^N) \right]
\]
(A11)
where \(\Delta_{c,\tau}\) is complementary volume to \(\Delta_c\), \(d\mathbf{R}^{N-1}_c = d\mathbf{R}_{2,\tau} \ldots d\mathbf{R}_{N,\tau}\) (similarly for \(d\mathbf{R}^{N-1}_c\)) and \(\mathbf{R}_i^N = \mathbf{R}_{i,\tau}, \ldots, \mathbf{R}_{2,\tau}\). Using Eq. (A10) and the expansion of the density \(\rho(\mathbf{R}_i^N)\) up to linear order (see Eq. (A5)) in the displacement from \(\mathbf{R}^N\),
\[
\rho(\mathbf{R}_i^N) = \rho(\mathbf{R}) + \sum_{j=1}^{N} \left[ \nabla \mathbf{R}_j \rho(\mathbf{R}^N) \right] \cdot \mathbf{v}(\mathbf{R}_j) d\tau,
\]
(A12)
the shape derivative of \(P^{(1)}(\Delta)\) becomes
\[
\frac{dP^{(1)}(\Delta)}{d\tau} =
\sum_{i=1}^{N} \int_{\Delta} d\mathbf{R}_i \left( \prod_{\ell \neq i} \int_{\Delta_c} d\mathbf{R}_\ell \right) \left[ \nabla \mathbf{R}_i \cdot \left( \rho(\mathbf{R}^N) \mathbf{v}(\mathbf{R}_i) \right) + \sum_{j \neq i} \nabla \mathbf{R}_j \cdot \left( \rho(\mathbf{R}^N) \mathbf{v}(\mathbf{R}_j) \right) \right].
\]
(A13)
The first term on the right-hand-side can be treated exactly as we have done in Eq. (A6), leading to
\[
\int_{\Delta} d\mathbf{R}_i \left( \prod_{\ell \neq i} \int_{\Delta_c} d\mathbf{R}_\ell \right) \nabla \mathbf{R}_i \cdot \rho(\mathbf{R}_i^N) \mathbf{v}(\mathbf{R}_i) = \int_{\partial\Delta} ds_i \mathbf{n}(\mathbf{R}_i) \cdot \mathbf{v}(\mathbf{R}_i) \left( \prod_{\ell \neq i} \int_{\Delta_c} d\mathbf{R}_\ell \right) \rho(\mathbf{R}_i^N)
\]
\[
= \int_{\partial\Delta} ds \mathbf{n}(\mathbf{R}) \cdot \mathbf{v}(\mathbf{R}) \int_{\Delta} d\mathbf{R}_i \delta(\mathbf{R}_i - \mathbf{R}) \left( \prod_{\ell \neq i} \int_{\Delta_c} d\mathbf{R}_\ell \right) \rho(\mathbf{R}_i^N).
\]
(A14)
The second line is obtained from the first by introducing a \( \delta \)-function in order to change the variable \( \mathbf{R}_i \) to \( \mathbf{R} \) in the surface integral. Notice now that the integral over the variable \( \mathbf{R}_i \) is performed over the whole space.

The second term in Eq. (A13) has still to be appropriately simplified. The \( N - 1 \) integrals of the second term in square brackets are transformed, one by one, in \( N - 1 \) surface integrals over \( \partial \Delta_c \), with \( \nabla_{\mathbf{R}_i} \) replaced by the normal vector to the surface, \( \mathbf{n}_c(\mathbf{R}_j) = -\mathbf{n}(\mathbf{R}_j) \) (when the volume \( \Delta \) changes in the direction indicated by \( \mathbf{n} \), the volume \( \Delta_c \) changes in the direction \( -\mathbf{n} \), since the boundaries of \( \Delta \) and \( \Delta_c \) are the same, being them complementary volumes to each other).

Therefore, we write explicitly each term of the sum over the index \( j \) in Eq. (A13).

\[
\int_{\Delta} d\mathbf{R}_i \left( \prod_{l \neq i}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \sum_{j \neq i}^{N} \nabla_{\mathbf{R}_j} \cdot \left( \rho(\mathbf{R}^N) \mathbf{v}(\mathbf{R}_j) \right) = \\
= \int_{\Delta} d\mathbf{R}_i \int_{\partial \Delta_c} ds_1 \left( \prod_{l \neq i, 1}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \mathbf{n}_c(\mathbf{R}_1) \cdot \mathbf{v}(\mathbf{R}_1) \rho(\mathbf{R}^N) \\
+ \int_{\Delta} d\mathbf{R}_i \int_{\partial \Delta_c} ds_2 \left( \prod_{l \neq i, 2}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \mathbf{n}_c(\mathbf{R}_2) \cdot \mathbf{v}(\mathbf{R}_2) \rho(\mathbf{R}^N) + \ldots
\]

obtaining

\[
\int_{\Delta} d\mathbf{R}_i \left( \prod_{l \neq i}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \sum_{j \neq i}^{N} \nabla_{\mathbf{R}_j} \cdot \left( \rho(\mathbf{R}^N) \mathbf{v}(\mathbf{R}_j) \right) = \\
\int_{\Delta} d\mathbf{R}_i \sum_{j \neq i}^{N} \int_{\partial \Delta_c} ds_j \left( \prod_{l \neq i, j}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \mathbf{n}_c(\mathbf{R}_j) \cdot \mathbf{v}(\mathbf{R}_j) \rho(\mathbf{R}^N). \quad \text{(A16)}
\]

We introduce now a \( \delta \)-function, in order to make a change of variable \( \mathbf{R}_j \rightarrow \mathbf{R} \) and to write the above integral in compact form, namely

\[
\int_{\Delta} d\mathbf{R}_i \left( \prod_{l \neq i}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \sum_{j \neq i}^{N} \nabla_{\mathbf{R}_j} \cdot \left( \rho(\mathbf{R}^N) \mathbf{v}(\mathbf{R}_j) \right) = \\
\int_{\Delta} d\mathbf{R}_i \sum_{j \neq i}^{N} \int_{\partial \Delta_c} ds \mathbf{n}_c(\mathbf{R}) \cdot \mathbf{v}(\mathbf{R}) \int d\mathbf{R}_j \delta(\mathbf{R}_j - \mathbf{R}) \left( \prod_{l \neq i, j}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \rho(\mathbf{R}^N) \quad \text{(A17)}
\]

\[
= -\int_{\Delta} d\mathbf{R}_i \sum_{j \neq i}^{N} \int_{\partial \Delta} ds \mathbf{n}(\mathbf{R}) \cdot \mathbf{v}(\mathbf{R}) \int d\mathbf{R}_j \delta(\mathbf{R}_j - \mathbf{R}) \left( \prod_{l \neq i, j}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \rho(\mathbf{R}^N). \quad \text{(A18)}
\]

From Eq. (A17) we have obtained Eq. (A18) by replacing \( \mathbf{n}_c \) with \( -\mathbf{n} \) and by using the property that the boundary of \( \Delta_c \) is the boundary of \( \Delta \). As previously observed for Eq. (A14), the integral over the variable \( \mathbf{R}_j \) is performed over the whole space.

Combining the results in Eqs. (A14) and (A18), we derive the expression for the shape derivative of the one-particle occupancy probability as

\[
\frac{dP^{(1)}(\Delta)}{d\tau} = \int_{\partial \Delta} ds \mathbf{n}(\mathbf{R}) \cdot \mathbf{v}(\mathbf{R}) \left\{ \sum_{i=1}^{N} \left[ \int d\mathbf{R}_i \delta(\mathbf{R}_i - \mathbf{R}) \left( \prod_{l \neq i}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \right] \right\} \\
- \int_{\Delta} d\mathbf{R}_i \sum_{j \neq i}^{N} \int d\mathbf{R}_j \delta(\mathbf{R}_j - \mathbf{R}) \left( \prod_{l \neq i, j}^{N} \int_{\Delta_c} d\mathbf{R}_l \right) \rho(\mathbf{R}^N). \quad \text{(A19)}
\]

The characteristic functions \( \Upsilon_{\Delta}(\mathbf{R}_j) \) and \( \Upsilon_{\Delta_c}(\mathbf{R}_j) = 1 - \Upsilon_{\Delta}(\mathbf{R}_j) \) are used in order to extend the integrals over \( \Delta \) and \( \Delta_c \) to the whole space. We then obtain

\[
\frac{dP^{(1)}(\Delta)}{d\tau} = \int_{\partial \Delta} ds \mathbf{n}(\mathbf{R}) \cdot \mathbf{v}(\mathbf{R}) f_{\Delta}(\mathbf{R}) \quad \text{(A21)}
\]
with
\[
    f_\Delta(R) = \left\langle \sum_{i=1}^{N} \left[ \delta(R_i - R) \prod_{l \neq i}^N (1 - \Psi_\Delta(R_l)) - \Psi_\Delta(R_i) \sum_{j \neq i,j}^N \delta(R_j - R) \prod_{l \neq i,j}^N (1 - \Psi_\Delta(R_l)) \right] \right\rangle. \tag{A22}
\]

If the arbitrary velocity field in Eq. (A21) is chosen to be \( v(R) = n(R)f_\Delta(R) \) as in Eq. (A8), the condition \( D_\Delta P^{(1)}(\Delta) = dP^{(1)}(\Delta)/d\tau \geq 0 \) is again automatically satisfied.

3. Level set function

In order to regularise the mathematical treatment of the characteristic functions, it is useful to introduce a family of regular functions collectively defined as level set functions. A level set function \( \phi(R, \tau) \) is defined by the property

\[
    \phi(R, \tau) \begin{cases} < 0, & R \in \Delta \\ = 0, & R \in \partial \Delta \\ > 0, & R \notin \Delta \end{cases} \quad \forall \tau, \tag{A23}
\]

such that \( \partial \Delta \) is identified as the set of zeros of \( \phi(R, \tau) \). The normal vector to the iso-surfaces of this smooth function is defined as

\[
    n(R) = \frac{\nabla \phi(R, \tau)}{|\nabla \phi(R, \tau)|} \tag{A24}
\]

and, in particular for the iso-surface \( \phi(R, \tau) = 0 \), \( n(R) \) is the normal vector to the boundary of \( \Delta \) that appears in Eq. (A21).

To define the evolution of the set, we ask that the level set function \( \phi(R, \tau) \) be such that its total time-derivative is 0. This results in the following evolution equation

\[
    0 = \frac{d\phi(R_\tau, \tau)}{d\tau} = \frac{\partial \phi(R_\tau, \tau)}{\partial \tau} + v(R_\tau) \cdot \nabla \phi(R_\tau, \tau), \tag{A25}
\]

where the velocity field \( v(R) \) at \( R_\tau \) will be chosen to be

\[
    v(R_\tau) = n(R_\tau) f_\Delta(R_\tau) = \frac{\nabla \phi(R_\tau, \tau)}{|\nabla \phi(R_\tau, \tau)|} f_\Delta(R_\tau). \tag{A26}
\]

Eq. (A25) guarantees that the iso-surface \( \phi(R_\tau, \tau) = 0 \) mimics the evolution of the boundary of \( \Delta \) according to the deformation law determined by imposing the condition \( D_\Delta P^{(1)}(\Delta) \geq 0 \) on the shape derivative of the one-particle occupancy probability. It is important to underline that imposing Eq. (A25) means that, at the boundary of \( \Delta \), the value of the level set function does not change in time. Therefore, we are able to identify at all times the set of points defining \( \partial \Delta_\tau \). Further, by imposing that Eq. (A25) is valid everywhere in space, we are obtaining the deformation law \( \Delta \rightarrow \Delta_\tau \) by evolving the auxiliary level set function. The advantages are the possibility of calculating the normal vector \( n(R) \), being \( \phi(R, \tau) \) a smooth function of \( R \) by construction, and of being able to identify at each time the domain \( \Delta_\tau \) (the characteristic function is constructed by knowing where \( \phi(R, \tau) \) is positive). We approximate the right-hand-side of Eq. (A25) by a finite difference

\[
    \phi(R_\tau, \tau + d\tau) = \phi(R_\tau, \tau) - d\tau f_\Delta(R_\tau)|\nabla \phi(R_\tau, \tau)|, \tag{A27}
\]

using Eq. (A24) for the unitary vector normal to the boundary of \( \Delta \) (or to the isosurface \( \phi(R_\tau, \tau) = 0 \)). This equation is then transformed in an algorithm that determines to evolution of the domains \( \Delta_i \), given an initial condition for the level set function corresponding to the choice of initial domains discussed in the text.

Appendix B: The ideal gas

The probability \( P^{(1)}(\Delta) \) can be analytically evaluated for the ideal gas. For a non-interacting system, \( P^{(1)}(\Delta) \) does not depend on the shape of the domain \( \Delta \) but only on its volume \( v \). Thus, there is total degeneracy in the shape when looking for an optimal domain because only the volume (and not the shape) of the MPD can be determined.
from the maximisation of the one-particle occupancy probability. This fact will reduce the calculation of the MPD to a trivial analysis exercise which we will give below.

Eq. (1) in this case becomes

$$P^{(1)}(\Delta) = \frac{1}{V^N} \left( \frac{N}{1} \right) \int_d R^1 \int_{\nu - v} dR^{N-1} = N \left( \frac{\nu}{V} \right) \left( 1 - \frac{\nu}{V} \right)^{N-1}$$

(B1)

where $V^N$ is the configurational partition function for the ideal gas. Here we have indicated the complementary volume of $\nu$ as $V - \nu$. The thermodynamic limit is now easily obtained by observing that

$$\lim_{N \to \infty} P^{(1)}(\Delta) = \left( \frac{\nu}{\nu_0} \right) \lim_{N \to \infty} \left( 1 - \frac{\nu}{\nu_0} \right)^{N-1} = \frac{\nu}{\nu_0} e^{-\frac{V}{\nu}}$$

(B2)

where $\nu_0 = V/N$ is kept constant and we have eliminated any dependence on $V$. It is natural to ask at this point what is the volume $\nu$ that maximises this probability and what is the value of this maximum probability. We can answer these questions by just differentiating Eq. (B2) with respect to $\nu$, since in this case the shape derivative reduces to the standard derivative. We find that $dP^{(1)}(\nu)/d\nu = 0$ if $\nu = \nu_0$. For the value of the probability at the optimal volume, we find $P^{(1)}(\nu_0) = e^{-1} = 0.37$.
[38] T. Strässle, A. M. Saitta, Y. L. Godec, G. Hamel, S. Klotz, J. S. Loveday, and R. J. Nelmes, Phys. Rev. Lett. 96, 067801 (2006).