On the Role of Non-Spherical Cavities in Short Length-Scale Density Fluctuations in Water

Gabriele Cesare Sosso,^{*,†} Sebastiano Caravati,[‡] Grant Rotskoff,[¶] Suriyanarayan Vaikuntanathan,[§] and Ali Hassanali^{*,||}

†Thomas Young Centre, London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, Gower Street London WC1E 6BT, United Kingdom ‡Department of Chemistry, University of Zurich, Winterhurerstrasse 190 CH-8057, Switzerland

¶Biophysics Graduate Group, University of California, Berkeley, CA 94720, USA §Department of Chemistry, The University of Chicago, IL 60637, USA

||Condensed Matter and Statistical Physics Section, The Abdus Salam International Centre for Theoretical Physics, I-34151 Trieste, Italy

E-mail: g.sosso@ucl.ac.uk; ahassana@ictp.it

Abstract

Density fluctuations in liquid water are at the heart of numerous phenomena associated with hydrophobic effects such as protein folding and the interaction between bio-molecules. One of the most fundamental processes in this regard, is the solvation of hydrophobic solutes in water. The vast majority of theoretical and numerical studies examine density fluctuations at the short length-scale focusing exclusively on spherical cavities. In this work, we use both first principles and classical molecular dynamics simulations to demonstrate that density fluctuations in liquid water can deviate significantly from the canonical spherical shapes. We show that regions of empty space are frequently characterized by exotic, highly asymmetric shapes that can be quite delocalized over the hydrogen bond network. Interestingly, density fluctuations of these shapes are characterized by Gaussian statistics with larger fluctuations. An important consequence of this is that the work required to create non spherical cavities can be substantially smaller than that of spheres. This feature is also qualitatively captured by the Lum-Chandler-Weeks theory. The scaling behavior of the free energy as a function of the volume at short length scales is also qualitatively different for the non-spherical entities. We also demonstrate that non spherical density fluctuations are important for accommodating the hydrophobic amino acid alanine and are thus likely to have significant implications when it comes to solvating highly asymmetrical species such as alkanes, polymers or bio-molecules.

1 Introduction

Hydrophobic effects and hydrophobic interactions have significant implications in determining the structure, dynamics, aggregation and subsequently the function of bio-molecules in aqueous environments .^{1–4} This is at the heart of important physical and chemical processes in biophysics such as protein folding and the aggregation of bio-molecules. One of the most fundamental and studied processes in this regard is the accommodation of hydrophobic solutes in water. From the point of view of thermodynamics, the first step of hydrophobic solvation involves determining the reversible work required to create an empty space or cavity in the hydrogen bond (HB) network. Understanding the microscopic and molecular ingredients of hydrophobic solvation has been the subject of numerous theoretical as well as experimental studies.^{5–11}

Over the last few decades, there have been several theoretical and computational studies examining the molecular signatures of hydrophobicity.^{6,12–16} An aspect of this problem that has consistently emerged is how solvent density fluctuations are altered in close vicinity to a hydrophobic surface.^{17–20} In particular, the statistical mechanics of cavity (a connected region of empty space within the water network) formation and how this changes as a function of the size of the hydrophobic solute, has been shown to be a good measure of hydrophobicity. At the same time, the properties of density fluctuations provide a microscopic link between the small and large length scale hydrophobic effects where in the latter regime the thermodynamics is driven by the surface energy of interface formation.^{7,20}

For the most part, cavities in water are treated as highly idealized spherical objects. In this context, quantities such as the excess chemical potential for inserting a cavity in water, as used in the information theory for hydrophobic effects 21 examine the number density fluctuations that occur within spherical regions. There have been recent studies examining the nature of density fluctuations (i.e. the geometrical and thermodynamical properties of regions of empty space forming within the water network) for non-spherical geometries such as cubes, cuboids and cylinders.^{22,23} The solvation thermodynamics of non-spherical shapes has also been explored with morphometric thermodynamics²⁴ and scaled-particle theory.²⁵ However, these shapes are still idealized as they do not capture the actual shapes that emerge from microscopic inhomogeneities and molecular roughness that characterizes regions of free space in the HB network. Sulimov and co-workers examined the cavitation free energies of organic molecules consisting of different sizes and shapes with molecular dynamic simulations and found that for systems with an effective radii less than 7 Å the computed free energies appeared to exhibit a linear dependence in the volume.²⁶ Recently Pettitt and co-workers have examined the effects of geometry and chemistry on hydrophobic solvation of hydrocarbon molecules and found that hydrophobic theories built on assuming a simple linear dependence of the free energy of solvation on the surface area or volume were not consistent with their findings from simulations.²⁷ This evidence calls attention to the need of furthering our understanding of hydrophobic interactions in order to improve the current hydrophobic models.

In this work, we re-visit the statistical properties of short length scale density fluctuations using molecular dynamics (MD) simulations of liquid water and show that the thermodynamics of the creation of realistic, non-spherical cavities in water is qualitatively different from that of spheres. In particular, we employ the Voronoi-Delaunay method²⁸ to characterize the regions of empty space within the water network. The resulting Voronoi-Delaunay voids (VDV) are characterized by a much richer diversity of morphologies when compared to the spherical cavities (SC) typically employed to investigate hydrophobicity at the molecular level. The VDV are highly non spherical and rather than being localized in one region of the network, can often be spread across several solvation shells. The VDV we identify are subsequently used to examine the statistics of number density fluctuations as has been done previously for SC. We find that VDV are essentially *less hydrophobic* than SC since the work required to create them is substantially smaller compared to a SC of the same volume. Interestingly, the density fluctuations associated with VDV are well described by Gaussian fluctuations. Furthermore, the difference between the density fluctuations of SC and VDV is also qualitatively captured by the Lum-Chandler-Weeks (LCW) Gaussian field theory.⁵ The microscopic fluctuations associated with hydrophobicity below the 1 nm length scale for the VDV are more nuanced than the picture built on the analysis of only SC leading to very different behavior in the transition from short-to-long length scale hydrophobic effects. The presence of asymmetric and exotic shapes in bulk water are a generic feature since we find that they are found not only in DFT-based *ab initio* water, but also in two other water models, TIP4P-Ew²⁹ and, as illustrated in the Supporting Information (SI), the MB-pol model.^{30–32} For processes that do not involve the dissociation of water, MB-pol is probably the most realistic model of molecular H_2O .

Understanding the density fluctuations of VDV has deep implications for water surrounding and in proximity of ions, rough molecular systems such as amino-acids, proteins and interfaces in general. Their branched and highly asymmetric character makes them good candidates for accommodating realistic small chemical systems. Here, we illustrate the possibility of these effects by examining the properties of the VDV that enclose the hydrophobic amino-acid alanine, in water. Similar to that of bulk water, the VDV distribution that engulfs alanine is highly asymmetric and requires less work to create than a SC of the same volume. The role of molecular roughness of both hydrophobic and polar organic molecules as manifested through the density fluctuations of VDV is an important future direction to explore in the area of hydrophobic effects.

2 Computational Details

2.1 Molecular Dynamics Simulations

Most of our analysis of both VDV and SC relies on the use of AIMD trajectories of periodic cubic liquid water boxes consisting of 128 water molecules. These simulations have recently been used to understand defect correlations in the HB network.³³ The reader is referred to the original paper for simulation details, however, we mention briefly that both the SC and

VDV involved an analysis of 4 independent AIMD simulations with a total simulation length of 64ps using the BLYP functional with Grimme's dispersion corrections. These simulations were thermostatted at 300 K. It should be mentioned that obtaining large statistical samples of VDV is computationally expensive and the smaller AIMD systems make this process tractable. We were obviously concerned about whether the small system sizes and short simulation times would bias the density fluctuations and particularly the VDV statistics associated with p_0 which is the probability to find zero waters within the cavity as described in detail later. Classical MD simulations of a large (4096 molecules) model of bulk water have thus also been performed via the MD code LAMMPS.³⁴ Re-constructing the VDV for this big system is computationally prohibitive and hence VDV sampled from the AIMD simulation were used to examine p_0 obtained in the larger simulations. As explained in the manuscript, obtaining quantitative agreement with previous studies requires the use of classical MD simulations.

For the classical MD, water molecules have been modeled using the TIP4P-Ew²⁹ rigid water model. The density of water has been enforced to reproduce that of the AIMD simulations used in this work (0.0334 mol/Å³), resulting in a cubic simulation box of length equal to 49.68 Å. The time step for the integration of the equations of motion and the coupling constant of the Nose-Hoover thermostat have been set to 2 and 1000 fs respectively. The system was equilibrated at 270 K for 10 ns since the melting temperature for the TIP4P-Ew water model is $T_m \sim 244$ K. The 10⁵ configurations used to accumulate statistics for both the SC and VDV (labeled SC_{T4} and VDV_{T4} for this system throughout the rest of the paper) have been extracted every 0.2 ps from a production run 20 ns long.

As alluded to in the introduction, we also performed some analysis on water trajectories at 300 K using the MB-pol potential.^{30–32} These simulations consist of 216 water molecules in a periodic cubic box of side length equal to 18.65 Å. Analysis of results obtained from MB-pol are shown in the SI (see e.g. Fig. S8).



Figure 1: Typical shape of a) small (~50-100 Å³) and b) large (~150-250 Å³) VDV as found within a 128-mol model of ab initio water at room temperature. While small VDV tend to be rather spherical and compact, larger ones often display rather complex, non-spherical shapes. Red dotted lines in panel b) highlight the bottlenecks (see text) within the large VDV. The inset in panel b) (top right) depicts a single water molecule on the same scale. The volume of the VDV is 88 and 249 Å³ in panels a) and b) respectively.

2.2 Voronoi Voids

Our goal in this work was to determine the realistic shape and volume of regions of empty space within the water network. One popular approach used when dealing with bubble nucleation and cavitation in general ^{35,36} is to use grid-based methods where the simulation box is partitioned into three-dimensional regular cells usually of cubic shape. Volume elements that do not contain any atoms within them are subsequently connected to obtain the regions of empty space in the system. Despite the usage of these structured³⁷ grids due to their computational convenience, these probes are not related to the topological features of the atomic/molecular network, and as such, they usually cannot deal accurately with different atomic/molecular species for example, characterized by different van der Waals radii. Most importantly, taking into account highly non-spherical cavities by means of structured grids characterized by uniform cells size is often a challenging task as the volume elements have

to be reduced down to a very small size, which in turn can lead to nonphysical pathways between different cavities.

For clarity, we remark that the analysis of the VDV we report here is distinct from evaluating the properties of Voronoi polyhedra of water which has been reported several times in the literature .^{38,39} In fact, our analysis explicitly accounts for the atomistic structure of water molecules since the Voronoi S-network is a generalization of the Voronoi-Delaunay tessellation for multicomponent (or *polydisperse*) systems, namely ensembles of atoms having different radii.

In this work, we adopt instead a flexible and robust approach based on the 3D Voronoi S-network proposed by Medvedev *et al.*^{28,40} as implemented in the VNP program.⁴¹ This methodology has been extensively employed to investigate cavities (or Voronoi-Delaunay voids, VDV) within a number of systems, including molecules in solutions⁴² and complex interfaces.⁴³ We have also recently applied this particular framework to elucidate the emergence of nano cavities in seminconducting glasses⁴⁴ and to characterize the molecular mechanism of proton diffusion in water.⁴⁵ The 3D Voronoi S-network is a clever approach that builds upon the actual topology of the system, resulting in an unstructured three dimensional grid capable of taking into account the complex, non spherical shapes of empty regions in a HB network.

Here we review some of the essential concepts and refer the reader to the works by Medvedev *et al.* for details. In a polydisperse system, atomic volumes are delimited by the *Voronoi S surfaces* (VS surfaces). A VS surface is defined as the geometric locus of points equidistant from the surfaces of two atomic spheres instead of from their centers. VS surfaces are no longer planes as for Voronoi polyhedra but hyperboloids and their intersection gives rise to curved edges (VS channels) and vertexes (VS vertexes). An interstitial sphere with radius R_I , tangent to neighbor atoms, is associated with each VS vertex while each VS channel is characterized by the radius R'_B of the largest test particle moving along that channel. The network of VS vertexes interlinked by VS channels can be used to distinguish interatomic voids and resolve connected cavities. In fact, a natural way of defining voids is through the value of the radius (R_P) of a test particle (probe) that can be put in the given void. Thus, voids are the parts of the interatomic space that are accessible for a given probe (i.e. $R_I \ge R_P$, R_I criterion). On the other hand, if a probe can move along an S-network bond (i.e. pass through a bottleneck) then both VS vertexes connected by that VS channel belong to the same cavity. Thus, if the extent of the bottleneck (represented by the radius R'_B of the largest test particle that would fit into it) is greater than a certain threshold R_B (which we refer to as *bottleneck radius* hereafter), the two regions are considered connected (i.e. $R'_B \ge R_B$, R_B criterion).

In our simulations, R_P has been set to 1.20 Å since the Van der Waals radii of the hydrogen and oxygen atoms is 1.52 and 1.20 Å respectively while R_B has been set to 1.1 Å. The choice of R_B and R_P is important as they influence the number, the size and even the morphology of the VDV. In a previous work⁴⁵ we have extensively investigated different choices of R_B and R_P leading to meaningful VDV characterized by bottlenecks of size comparable to that of a water molecule and quite often even larger (see for example Fig. 1b). This is because the bottlenecks have to be larger than $R_B \cdot 2 = 2.2$ Å, an extent large enough to accomodate the hydrogen-hydrogen distance for a water and methane molecule, which is of the order of 1.65 and 1.75 Å respectively. In addition, as in Ref.,⁴⁵ we have checked that our results are consistent within a reasonable interval of values for both R_B and R_P . In particular, we have chosen R_B and R_P carefully to avoid very large volumes of empty spaces and very exotic VDV networks characterized by several bottlenecks, possibly percolating through the whole of the water network. As a result, the differences between VDV and SC reported in this work are likely to be even more pronounced in reality.

To compute the volume of voids, the space encompassed by the Voronoi S surfaces of the probe particles is calculated numerically through a local mesh, dense enough to minimize the residual below a threshold. Such voids are then eventually merged if they belong to the same cavity domain, according the the R_B criterion, giving rise to an arbitrarily shaped cavity.

2.3 Cavity and Void Statistics

One of the central quantities discussed in this work is the probability P(n) of observing n solvent centers (oxygens atoms in this case) within a SC or a VDV of a given volume. Such a probability has been obtained for a given SC or VDV by averaging n over 10^5 configurations coming from either ab initio or classical simulations. For each configuration, we have positioned the SC or VDV under analysis at the center of the simulation box and subsequently counting the number of oxygens that can be found within it. In this way, we have constructed the P(n) reported in Fig. 3. While this process is straightforward for the SC as has been done in several previous studies,²¹ handling the VDV requires special care. There is no ambiguity in the calculation of the VDV volume, as the geometrical definition of the VDV is unique once the system topology as well as the R_B and R_P parameters have been specified.

In order to probe whether a solvent center lies within the VDV or not, we have constructed for each VDV a corresponding 3D polyhedron, taking advantage of the alpha shape⁴⁶ construction available in MATLAB.⁴⁷ Such an object leads to a very fast computation of P(n), as it allows to establish quickly whether a point (in our case, an oxygen atom) lies or not within the alpha shape, and thus within the VDV. The alpha shape construction for each VDV is not unique and depends on the alpha radius R_{α} , which in turn controls the level of detail of the alpha shape, or in other words how tightly the latter fits with respect to the original set of points (in our case, the vertexes of the Delaunay simplexes involved in the VDV). Thus, the choice of R_{α} affects the volume as well as the surface area of the resulting alpha shapes, that is, of the 3D polygons by which we represent the VDV in order to compute P(n). Hence, for each VDV we have determined R_{α} so that the volume of the corresponding alpha shape is consistent with the volume of the actual VDV within an accuracy of $+/- 5\text{Å}^3$. We have also checked that different values of R_{α} lead to very similar results in terms of the predicted surface area, introducing an uncertainty of - at most - 5%, far too small to have any effect on the results reported in this work.

In order to extract the excess chemical potential $\Delta \mu$ for SC/VDV formation, we have fitted each P(n) with a normal distribution $P_{Gauss}(n)$:

$$\Delta \mu = -k_B T \ln P_{Gauss}(n=0) \tag{1}$$

Here, k_B is the Boltzmann constant, T is the temperature, and $P_{Gauss}(n = 0)$ is the value of P(n) for n = 0. Consistent with previous theoretical and numerical studies, the P(n)distributions for the SC are well described by Gaussian fluctuations. Perhaps more surprising, is that the larger asymmetric VDV are also characterized by Gaussian statistics.

3 Results and Discussion

3.1 Voronoi-Delaunay Voids (VDV) vs Spherical Cavities (SC): Structural Properties

In order to build our intuition on the results and discussions to follow, we begin by discussing the structural properties of the regions of empty space we are going to consider, described either in terms of spherical cavities (SC) or Voronoi-Delaunay voids (VDV). To this end, we illustrate in Fig. 1 two different VDV obtained from an ab initio molecular dynamics (AIMD) simulation of liquid water. Details of the MD simulations and the procedure employed to extract the VDV from the water network are discussed in Section 2. Note that while the surface of the VDV has no analytic representation, it provides a convenient and robust numerical way to quantify the space enclosed by the VDV. Small (50-100 Å³) VDV such as the one depicted in Fig. 1a are typically rather spherical and compact. In contrast, large (150-250 Å³) VDV such as the one shown in Fig. 1b can be characterized by quite exotic shapes that deviates significantly from spherical symmetry. These morphologies emerge



Figure 2: a) Cumulative distribution function (CDF) of the volume of VDV and SC within the 128-mol model of ab initio water at room temperature. b) Before-After plot for the VDV volume (left stack) and the volume of the largest SC one can build by sitting on the baricenter of the correspondent VDV (right stack). Points and segments for which VDV volume is smaller than SC_{VDV} volume are depicted in blue, while points and segments for which VDV volume is larger than SC_{VDV} volume are depicted in red. c) Asphericity of VDV as a function of volume. Reference values for different regular shapes are also reported: SC, TH, CUB and L_{CUB} refer to sphere, tetrahedron, cube and asymmetrical cuboid. d) Surface area S_A as a function of volume for SC and VDV.

from the spontaneous fluctuations in the HB network, and in some cases they extend across several solvation shells. While some of these large regions of empty space encompass just one convex polyhedron (see Fig. 1a), in many situations their branched shapes originate from the existence of bottlenecks (see Fig. 1b). It should be noted that in most cases these channels connecting two regions of empty space are wide enough to accommodate at least one water molecule. One can imagine how such branched VDV could provide the right ingredients to accommodate elongated shapes of molecules such as branched alkane chains and small peptides.

The VDV shown in Fig. 1b already illustrate that there are likely to be regions of empty space that are much larger in volume and surface area than those one would predict looking exclusively at SC within the same water network. To see this more quantitatively, we show in Fig. 2a the cumulative distribution function (CDF) of the volume of SC and VDV within the water network that are observed from spontaneous thermal fluctuations at room temperature. While the probability of finding SC of volume > 100 Å³ is basically zero, fat tails in the CDF for the VDV indicates the presence of sizable regions of empty space up to 200-250 Å³. For the SC, a volume of 250 Å³ corresponds to a SC with radius of about 4 Å. Needless to say, larger SC can possibly form in water, but cannot be found within the relatively short times scale of our ab initio molecular dynamics (AIMD) simulation.

Probing the density fluctuations in water focusing exclusively on the SC does not capture the complexity of the underlying network originating from the molecular roughness and inhomogeneities in the 3D-structure of the hydrogen bond network. One way to see this is to look for SC that can form within non-spherical VDV. This is seen in Fig. 2b where we report the volume of the VDV (left stack) against the volume of the largest SC (SC_{VDV}, right stack) that can be built by sitting on the center of mass of the corresponding VDV. Firstly, we see that small (~50-100 Å³) VDV are slightly smaller (dotted blue segments pointing upward from the VDV to SC_{VDV}) than the corresponding SC_{VDV}. This is because the surface of a SC exactly touches the closest atom with respect to its center of mass, while the Voronoi construction upon which the VDV are built implies that some space around the atoms is excluded according to the vdW radius of each atomic species. This feature holds for small VDV characterized by a fairly spherical shape. On the other hand, larger (~150-250 Å³) VDV possess a much larger volume (solid red segments pointing downward) with respect to the corresponding SC_{VDV}. This is because large VDV are characterized by branched, non spherical shapes including bottlenecks which cannot be captured by SC. Yet another probe into the complexity of the shape of the VDV is their asphericity which can be quantified using the following expression:⁴⁸

$$\eta = \frac{S_A^3}{36\pi V^2} \tag{2}$$

where S_A is the surface area of the VDV and V its volume. A scatter plot of η against the volume of the VDV is reported in Fig. 2c together with the reference values for a regular sphere (SC), a regular tetrahedron (TH), a cube (CUB) and a very asymmetrical cuboid (L_{CUB}) characterized by an edge ten times larger than the other two sides. One can clearly see that shapes of larger VDV in the HB network deviate significantly from idealized objects. In particular, a number of large VDV feature an asphericity value far greater than the most common regular polyhedra, including the elongated shape of L_{CUB} . This is due to the fact that the realistic regions of free space in the hydrogen bond network as probed by the VDV consist of highly corrugated objects characterized by a very large surface area as shown in Fig. 2d. Details about the calculation of the surface area of the VDV are included in Section 2.

In summary, our results up to this point show that the creation of the asymmetric and aspherical voids is a feature that is present in both DFT water and also the classical potential TIP4P/EW. As indicated earlier, we also repeated some of our analysis of the cavities using MB-pol which indicates that the formation of the VDV is a generic property that is independent of the type of water model.

4 VDV vs SC: Density Fluctuations and Thermodynamics of Solvation

As alluded to earlier, one of the most important signatures and studied properties for understanding hydrophobicity are density fluctuations in bulk water²¹ and near hydrophobic interfaces.⁴⁹ In this context, one often looks at the reversible work $\Delta \mu$ required to create a region of empty space within the water network. This free energy cost, often referred to as the *excess chemical potential* is related to the probability P(n) of finding n solvent centers (in our case: oxygen atoms) within that particular empty region via the following relation:

$$\Delta \mu = -k_B T \ln P(n=0), \tag{3}$$

where k_B is the Boltzmann constant, T is the temperature of interest and P(n = 0) is the probability of finding zero solvent centers within the cavity. Further details can be found in the Computational Details section. Given the abundance of highly non-spherical VDV with different shapes in the HB network, we now move on to understanding the thermodynamics of their creation.

4.1 Gaussian Fluctuations for VDV

As described in more detail in the Computational Details section, constructing the VDV is quite computationally intensive and hence it is only feasible to extract these from the smaller AIMD simulations. Obvious concerns when using AIMD are the short simulation times (~10² ps) as well as the small system sizes (128 water molecules), for converging properties such as the P(n). For this reason, we have computed the P(n) by probing the density fluctuations from classical MD simulations as well. The classical simulations allow for the accumulation of statistics over ~10⁴ ps for large (4096 water molecules) simulation boxes. Indeed, while AIMD and classical MD provide the same qualitative picture (as shown in Fig. 3c), we will see that the classical simulations provide a description of the thermodynamics of solvation for SC in quantitative agreement with previous studies .²¹ We thus report in Fig. 3) the P(n)obtained from the classical simulations using VDV that were extracted from the AIMD simulations. More details on this can be found in Section 2.

The P(n) shown in Fig.3 show that the density fluctuations associated with VDV are



Figure 3: Probability P(n) of observing *n* solvent centers (n oxygens) within VDV (VDV_{T4}) and a corresponding SC (SC_{T4}) of the same volume. Statistics has been accumulated over 10^5 configurations of a 4096-mol model of TIP4P/Ew water. Panels a) b) and c) report P(n) for selected VDV of 46, 97 and 198 Å³ respectively, together with insets showing the VDV shape as approximated via the alphaShape construction (see the Computational Details section). We also report in panel c) (right side) the P(n) obtained for the same VDV and SC (VDV and SC respectively) but calculated using 10^5 configurations of a 128-mol model of BLYP water (see Section 2).

characterized by softer tails to both lower and higher n when compared to the P(n) obtained for an SC of the same volume. This implies that P(n = 0) for the VDV is larger than what we observe for SC. In other words, Eq. 3 tells us that the free energy cost needed to create a VDV of a given size can be substantially smaller than that required to form a SC. It is worth stressing that this particular trend holds for the entire range of VDV volumes examined. In the SI (see Fig. S1 and Fig. S2) we report the P(n) for VDV and SC obtained from AIMD and classical simulations for some more SC and VDV.

Having assessed the properties of P(n) for SC and VDV, we now turn our attention to the behavior of $\Delta \mu$ as a function of an effective length scale. The LCW theory predicts a characteristic length-scale behavior to the excess chemical potential for hydrophobic hydration. In particular, the theory predicts that at short length scales, the chemical potential is roughly proportional to the volume for spherical cavities. This feature is also captured by information theory and numerical simulations by Hummer and co-workers.²¹ At longer length-scales, $\Delta \mu$ is proportional instead to the surface area.⁶

In order to construct a similar analysis for the VDV, we need to provide a measure for its characteristic length. For the sake of simplicity, we define the length-scale of the VDV as the radius required to generate a SC with the same volume. As we will see later, this *effective radius* underestimates how delocalized the VDV are over the network. This feature will be illustrated more explicitly later in the text. Fig. 4a shows $\Delta \mu$ obtained from both the AIMD and classical simulations. Interestingly, the excess chemical potential characteristic of VDV is consistently lower than that of the SC with the same volume - the work required to create a VDV in this short-length scale regime is lower by about 10 – 15kJ/mol. This feature is independent of the system size and simulation time since it is found in both AIMD and classical simulations. As mentioned earlier, our results from the classical simulations are in better agreement with the an earlier study (see data of Hummer at al. shown in Fig 4a).

Our earlier results show that the density fluctuations associated with SC or VDV are well characterized by Gaussian statistics. Within this assumption, one can write⁵⁰ an analytic expression for $\Delta \mu$ as:

$$\Delta \mu \sim k_B T \cdot \frac{\rho_w^2 V^2}{2\chi_V} + k_B T \cdot \frac{\ln(2\pi\chi_V)}{2} \tag{4}$$

where ρ_w , V and χ_V are the water density, the volume of the VDV/SC and the variance of P(n) for that particular VDV/SC respectively. The second term of Eq. 4 depends logarithmically on χ_V , so that it is often neglected. In the SI (Fig. S4a) we show that indeed the inclusion of such a term does not change the qualitative trend that sees VDV having a substantially lower $\Delta \mu$ if compared to SC of the same volume. Gaussian statistics of the P(n) as observed for both the SC and VDV rationalizes the trends observed in $\Delta \mu$ shown in Fig. 4. For SC, $\chi_V \sim V$,⁷ so that $\Delta \mu \sim k_B T \cdot \rho_w^2 V$. However, for the VDV examined here $\chi_V \sim V^{\alpha}$ with $\alpha = 1.07 \pm 0.01$ which in turn leads to a lower $\Delta \mu$ for VDV compared to that of the corresponding SC. Although the change in the value of alpha is small, we will see later that the difference in the behavior of $\Delta \mu/S_A$ for SC and VDV can be rationalized



Figure 4: a) Excess chemical potential $\Delta \mu$ as a function the radius r of SC and VDV. For VDV, r is defined as the radius of a sphere with exactly the same volume as the VDV. Note however that $\Delta \mu$ has been obtained from the P(n) calculated using actual VDV with shapes like those depicted in Fig. 1 and Fig. 3. We also show $\Delta \mu$ for the same SC and VDV evaluated within 10⁵ configurations of a 16384-mol model of TIP4P/Ew water (SC_{T4} and VDV_{T4}, see text). The excess chemical potential for a large VDV found within the TIP4P/Ew model and its corresponding SC are labeled as VDV_{LT4} and SC_{LT4} (see text for details). Data for SC reported in the work of Hummer et al.²¹ are also reported. b) Excess chemical potential divided by surface area as a function of volume for SC and VDV. Data for the same SC and VDV evaluated within 10⁵ configurations of a 16384-mol model of a large VDV found within the same SC and VDV evaluated within 10⁵ configurations of a 16384-mol model of SC and VDV. Data for the same SC and VDV evaluated within 10⁵ configurations of a 16384-mol model of TIP4P/Ew water (SC_{T4} and VDV_{T4}, see text) are also reported, together with the result obtained for a large VDV found within the TIP4P/Ew model and its corresponding SC, labeled as VDV_{LT4} and SC_{LT4} . The subscript LT4 correspond to the largest VDV that was found by performing the Voronoi analysis on the TIP4P/Ew simulations.

by the subtle differences in the exponents such as α within the Gaussian theory. See the SI for more details on how the χ_V scales with the volume for SC and VDV (see Fig. S4b).

4.2 VDV vs SC: Behavior of $\Delta \mu / S_A$

The preceding analysis shows that the thermodynamics of VDV formation is quite different from that of SC. From an intuitive point of view, one way to rationalize this behavior is that the larger amounts of free space associated with VDV originate from the inherent 3D-architecture of the HB network which naturally creates regions of depleted density in water, as we shall see later. This creates small vapor-like tunnels or bottlenecks as seen in Fig. 1b which result from connected smaller regions of depleted density. On the other hand, the construction of large spheres requires a much larger localized depletion of density. To appreciate the effect of the differences between VDV and SC on the length-scale dependence to the hydrophobic effect, we recall that the surface area of the VDV exhibits a much stronger volume dependence than that showed by the SC as reported in Fig. 2d. For instance, VDV of about 200 Å³ can have a surface area twice as large as the SC of the same volume. This feature has critical consequences with respect to the excess chemical potential per unit area, $\Delta \mu/S_A$, reported for both VDV and SC in Fig. 4b. In the case of SC, the $\Delta \mu/S_A$ that has to be paid in order to create larger regions of empty space in the water network increases linearly with the radius of the SC consistent with previous observations from information theory and molecular simulations.²¹ In contrast, despite having a much larger surface area, creating large VDV costs less energy and $\Delta \mu/S_A$ for the VDV is basically constant on the length-scale quantified by their effective-radii.

Without employing any sophisticated sampling methods like those developed by Patel and co-workers^{22,23} to probe larger length-scales of realistic VDV, we cannot conclusively say how $\Delta \mu/S_A$ approaches the macroscopic surface tension. However, the value of $\Delta \mu/S_A$ we obtain for VDV is ~6 kJ mol⁻¹/nm⁻² which is roughly 15% of the macroscopic surface tension of the air-vapor interface and 26% of the *effective* surface tension reported for idealized shapes by Patel and co-workers.²² It is clear from the preceding analysis that the short-length scale behavior of hydrophobic effects for VDV is quite different from that of SC and that the details of the shape and geometry of hydrophobic cavities will play an important role in how abruptly the short-to-long hydrophobic behavior occurs.

4.3 VDV vs SC: LCW Theory

Molecular simulations by Hummer and co-workers have demonstrated that small length scale density fluctuations in liquid water are remarkably Gaussian.²¹ Indeed, the results reported in Fig. 3 indicate that the P(n) for both SC and VDV can be well described by Gaussian fluctuations. According to the LCW framework, density fluctuations at long and



Figure 5: Probability P(n) of observing n solvent centers (n oxygens) within VDV (VDV_{T4}) and a corresponding SC (SC_{T4}) of the same volume. Statistics has been accumulated over 10^5 configurations of a 4096-mol model of TIP4P/Ew water. The results obtained within the Gaussian approximation (see text) for VDV (VDV,G) and SC (SC,G) are also reported as dotted lines.

short wavelengths, corresponding to the emergence of large and small cavities within the water network respectively, are treated separately: while a purely Gaussian description is sufficient to deal with small (i.e. sub-nanometer scale) cavities, the thermodynamics of larger regions of empty space is dominated by the macroscopic surface tension. As such, the large distortions of the density field associated with the occurrence of such large cavities are typically resolved by coarse-graining the water network on a lattice.²³ This approach requires only the experimentally determined bulk pair correlation function and macroscopic surface tension as inputs, and is capable of dealing with density fluctuations involving very different length scales.⁵¹

In order to understand whether LCW theory is capable of describing the highly nonspherical density fluctuations resulting in the formation of the VDV, we compare in Fig. 5 the P(n) for selected SC and VDV we have obtained from the Gaussian fits of the atomistic simulations (SC/VDV_{T4}, obtained in the same fashion to those reported in Fig. 3), with those calculated from LCW theory using only the Gaussian contribution (labeled SC/VDV,G. See Refs. 23 and 51 for further details). Fatter tails in the P(n) for VDV,G with respect to SC,G can be observed for all the SC/VDV considered, in agreement with the SC/VDV_{T4} data. Thus, LCW theory confirms that the free energy cost needed to form a VDV of a given volume can be significantly lower than that required to create a SC of the same volume. Thus the difference in thermodynamics observed for SC and VDV is essentially encoded by the fluctuations underlying the pair correlation function of liquid water.

We note that although LCW theory predicts fatter tails of the VDV distributions compared to SC, the agreement with the numerical simulations is not quantitative - in fact, the errors appear to be slightly larger for the SC than the VDV. The origin of these differences is associated with errors originating from numerical integration and discretization $.^{23,51}$ In earlier studies, it was also shown that the long-wavelength lattice fluctuations were important for spheres of radii larger than 0.5 nm. In order to assess whether lattice fluctuations improve the characterization of the P(n) for the larger more delocalized VDV, distributions were constructed including the lattice contribution. Although there is some indication that doing so improves the P(n) for the VDV, lattice fluctuations do introduce large errors due to the small length scales of the SC/VDV considered in this work and hence we focus on only qualitative trends (see the SI for more details).

5 The HB Network in Proximity of VDV

Our analysis alluded to the delocalized nature of the VDV across the HB network. It is interesting to examine whether these features result in any significant disruption of the HB network. One way to probe this is to examine certain topological properties in water such as the deviation of the water molecules enveloping a VDV from local tetrahedrality. In order to quantify this, we examined both the in-degree (the number of HB being accepted by a water molecule) and the out-degree (the number of HB being donated) of water molecules that are within 2 Å of the vertices of the Voronoi-polyhedra enclosing the VDV. A similar analysis was recently performed on bulk water looking at defect correlations in the network.³³ In this study, it was found that independent of the choice of the water model, most water molecules in bulk water accepted and donated two HB and that thermal fluctuations result in defects

that cluster together. For the water molecules enveloping the VDV, the notion of their origin emerging from natural fluctuations is reinforced. In fact, the distributions of the in and out degrees show that most of these waters are tetrahedral - there is no signature of an increase in the number of defect water molecules such as over or under-coordinated water molecules compared to bulk water (see Fig. S5 in the SI).

The presence of the large VDV in the network can also be probed through the lens of other topological parameters. In liquid water, it is well known that local tetrahedrality results in the presence of closed rings having specific directional correlations leading subsequently to medium-range water wires.⁵² The extended VDV can be viewed as resulting from the specific correlations of the arrangement of rings with different sizes in the 3D-network. In order to examine this in more detail, we studied the radial density distributions between the center of mass of rings enveloping the VDV. As we did earlier for the defect distributions, rings were identified for water molecules within 2 Å of the vertices of the Voronoi-polyhedra. Fig. 6a shows the radial densities obtained between all 5, 6 or 7 membered rings enveloping large VDV indicating correlations extending up to ~ 8 Å. The 3D stacking of the closed rings and their corresponding fluctuations are not unique or exclusive to the immediate environment around a VDV. To see this more clearly, Fig. 6b illustrates the radial densities between the same sized rings, now averaged over all the rings in the water network. The distributions are indeed quite similar and reinforce the notion that the VDV are formed by the inherent correlations of the 3D HB network. The 3D-stacking of closed closed rings in the network essentially provides an intuitive way to understand how the VDV form.

These findings suggest that the creation of even significantly large ($\sim 250 \text{ Å}^3$) VDV does not require a sizable disruption of the HB network. It would be interesting in future studies to examine how the formation of larger idealized shapes such as spheres or cuboids alter properties such as the distribution of rings as well as local topological defects associated with the coordination of water molecules.



Figure 6: Radial density distribution between the center of mass of 5-, 6- and 7-membered a) closed rings enveloping the VDV (depicted in the central inset) and b) closed rings in the whole of the water network. The solid red, dotted blue and dashed green curves correspond to the 5-5, 6-6 and 7-M radial distributions.

6 VDV vs SC: Solvation of Hydrophobic Solutes

Understanding the role of density fluctuations in the creation of VDV around hydrophobic solutes and proteins is beyond the scope of the current study. However, as hydrophobic interactions are of paramount importance in important physical and chemical processes in biophysics such as protein folding and the aggregation of bio-molecules, we wanted to assess the role of non-spherical VDV in solvating a realistic hydrophobic molecule. In particular, the comparison of $\Delta \mu$ for VDV vs SC naturally occurring in bulk water (see Sec. 4.1 and e.g. Fig. 4), shows that the excluded volume contribution to the solvation energy is very sensitive to the shape. Here we focused on the hydrophobic amino-acid alanine - using AIMD simulations of a single alanine residue in water, we determined the volume and surface area of the VDV in the water network identified after the residue is removed. The VDV identified was then used to examine whether it would naturally emerge from fluctuations in bulk water, as we did before, as well as to understand how the thermodynamics of their formation would compare with an SC of the same volume.

The VDV volume distribution peaks at about 100\AA^3 with a rather broad distribution ranging between $60 - 200\text{\AA}^3$ (see SI Fig. S6a). Similar to those in bulk water, the VDV enveloping alanine are highly asymmetric with much more exotic shapes than that inferred



Figure 7: Excess chemical potential for a randomly selected subset of the VDV (and the corresponding SC of the same volume) created by removing the alanine molecule from the water network. Results are averaged over $\sim 10^3$ configurations. The inset depicts a frequent situation where the presence of the alanine molecule produces a VDV that is substantially larger than the size of the molecule itself.

from SC. Based on using the volume, surface area and asphericity measures (reported in the SI, see Fig. S6b and Fig. S6c), these VDV fall within the range of those naturally occurring from thermal fluctuations in bulk water. It is indeed interesting to see that the natural fluctuations of the HB network in bulk water, such as the one depicted in the inset of Fig. 7, create the appropriate excluded volume to accommodate the hydrophobic amino acid alanine.

Similar to what was shown for bulk water, the excess chemical potential for the VDV is smaller than the SC by about 5-10 kJ/mol, as illustrated in Fig. 7a. Note that $\Delta \mu$ has been obtained using the same set of 10^5 configurations of a 4096-mol model of TIP4P/Ew water employed for the P(n) reported in Fig. 3 (VDV/SC_{T4}). Fig. 7b shows $\Delta \mu/S_A$ for the VDV and SC - once again, the scaling behavior is very different as observed earlier for the cavities in bulk water. For the VDV, $\Delta \mu/S_A$ is essentially flat over the volume range explored and takes on a numerical value very close to that observed earlier in bulk water, which is a small fraction of the macroscopic surface tension of the air-water interface. Thus, intrinsic molecular roughness of the realistic shapes that enclose the alanine residue increases the importance of surface area in hydrophobic solvation at short length scales. Due to the small simulation box sizes used to analyze these effects, we cannot make quantitative statements about the VDV in the water network outside of the amino acid. However, an examination of the fluctuations around alanine shows that there are rather large asymmetric VDV regions extending from the hydrophobic moiety as seen in the inset of Fig. 7. This evidence is far for being trivial: the fact that even a small hydrophobic molecule such as the alanine residue considered here is capable of originate regions of empty space extending well beyond the neighboring water molecules suggests that larger objects such as proteins can substantially alter the topology and the thermodynamics of density fluctuations in liquid water.

7 Discussion and Conclusions

One of the most important cornerstones of hydrophobic effects is its sensitivity to the underlying length scale of the associated hydrophobic solute. At short length scales, a combination of both analytic theories and numerical simulations have shown that Gaussian statistics accurately characterize density fluctuations. On the other hand at larger length scales, the thermodynamics of hydrophobic solvation is driven by interface formation. The vast majority of theoretical studies in this area focus on the formation of idealized spherically SC. The work ($\Delta \mu$) required to create a SC grows as the volume and hence $\Delta \mu/S_A$ grows as the radius of a sphere (see Figure 2 of Ref. 7).

In this work, we re-visited the picture that has been constructed for the behavior of short length scale density fluctuations built upon focusing exclusively on SC. The shape of hydrophobic solutes has been shown to be important for example in anion binding in the context of understanding Hofmeister chaotropes⁵³ as well as in the solvation of hydrocarbons²⁷ and other organic molecules.²⁶ A key message of this work is that the molecular shape of voids in water has important implications on the physics of hydrophobic solvation. In particular, we extract an exotic mixture of VDV that are formed by natural fluctuations of the HB network using a Voronoi analysis. The regions of excluded volume in the water network are highly asymmetric and can extend over rather large regions of the HB network. The molecular origins of these features are shown to come from inherent features of the 3D network of water such as the stacking of closed rings across several hydration shells. Density fluctuations of the VDV are well characterized by Gaussian statistics. An interesting consequence of this is that the variance associated with these fluctuations is larger for VDV compared to SC of the same volume. Thus, despite having a much larger surface area than spheres of the same volume, the work required to create VDV is much smaller.

The scaling behavior of $\Delta \mu/S_A$ is qualitatively different for the VDV compared to SC. On the length scales currently probed from spontaneous thermal fluctuations, $\Delta \mu/S_A$ is effectively constant. The origin of this behavior appears to be well explained by the Gaussian theory: the larger width of the P(n) for VDV results in $\Delta \mu \sim k_B T \cdot \rho_w^2 V^\beta$ with $\beta < 1$. At the same time, the surface area of the VDV also has a non-trivial volume dependence. In particular for SC, the surface area scales exactly as $V^{2/3}$, while for the VDV, our numerical simulations suggest that the surface area grows as $V^{0.82\pm0.01}$ (see the SI for more details - Fig. S4c). Subsequently, $\Delta \mu / S_A$ grows as V^{1/3} and V^{0.10±0.014} for SC and VDV respectively (see Fig. S5a and b). As indicated earlier, the exponents we extract here provide a qualitative way to rationalize the difference in the thermodynamics of scaling behavior of SC vs VDV within using the Gaussian theory. The physics of short length-scale density fluctuations in water are clearly different for SC and VDV. Given that the VDV we examine represent those that are most easily formed, our results probe the lowest boundary in the space of $\Delta \mu/S_A$ and volume or effective radius. It is likely that there will exist other types of VDV covering the thermodynamic region between SC and the VDV. Our results also pose interesting challenges for theory to predict the solvation of hydrophobic solutes that go beyond exclusively concave or convex shapes. In this work, we focused on density fluctuations at ambient conditions only - it is clear that exploring the properties of VDV across the phase diagram of water would be an important and interesting extension of this study in the future.

Our focus in this work is on the characterization of static properties of the VDV. An obvious question of interest is on the lifetime of the VDV. Qualitatively, based on visual inspection, small VDV of volume $\sim 50-80$ Å³ survive for less than a ps. However, larger VDV of volume greater than 150 Å³ can survive for timescales of up to 5 ps. During these time windows, the shape and volume of the VDV will obviously change. It will be interesting in the future to come up with more quantitative measures of the lifetime of VDV of different size and shape and perhaps more interestingly, develop experimental probes of SC and VDV in water. Besides its obvious interest in fundamental chemical physics of hydrophobic effects in bulk water, understanding the properties of VDV has deep implications on the solvation of solutes in general. For instance, we have showed in a previous work⁴⁵ that proton transfer in liquid water can be promoted by the occurrence of particular VDV,

capable of facilitating an umbrella-like flip of the hydronium. In turn, this peculiar inversion mode of the charged species alters the free energy barrier for proton transfer. While it is clear that providing a detailed understanding of the role of VDV in the solvation of realistic amino acids as well as in processes such as protein folding and ion solvation is beyond the scope of this work, we attempted to shed some light on this for the hydrophobic amino acid alanine. Interestingly, the natural fluctuations of the HB network create VDV that would accommodate the excluded volume region needed for alanine. It will be interesting to examine these features for other amino acids as well as molecules with different types of branching and roughness. Furthermore, it will also be interesting in the future to re-visit the analysis of density fluctuations around proteins and near hydrophobic and hydrophilic interfaces that have typically been probed using idealized spheres.⁴⁹

Acknowledgement

The authors are grateful to M. Ceriotti for providing the MB-pol molecular dynamics trajectory discussed in the SM. G.S thanks P.Pedevilla, J. Chen and G. Brandenburg for commenting on the original manuscript. G.S. also acknowledges the use of the UCL GRACE High Performance Computing Facility and of the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk) via the Materials Chemistry Consortium through the EPSRC grant number EP/L000202.

Supplementary Material

We provide supplementary information about (i.) The probability P(n) of finding n solvent centers within spherical cavities (SC) and Voronoi-Delaunay Voids (VDV); (ii.) The lattice model results with respect to SC and VDV; (iii.) The scaling exponent of the excess free energy of solvation as a function of volume and surface area; (iv.) The structural properties of the water network in the neighborhood of the VDV; (v.) The statistics of VDV as obtained for the solvation of alanine in water; (vi.) VDV and SC in MB-pol water.

References

- Blokzijl, W.; Engberts, J. B. F. N. Hydrophobic Effects. Opinions and Facts. Angew. Chem. Int. Ed. (English) 1993, 32, 1545–1579.
- (2) Pratt, L. R. Theory of Hydrophobic Effects. Annu. Rev. Phys. Chem. 1985, 36, 433–449.
- (3) Pratt, L. R.; Pohorille, A. Hydrophobic Effects and Modeling of Biophysical Aqueous Solution Interfaces. *Chem. Rev.* 2002, 102, 2671–2692.
- (4) Ben-Amotz, D. Hydrophobic Ambivalence: Teetering on the Edge of Randomness. J. Phys. Chem. Lett. 2015, 6, 1696–1701.
- (5) Lum, K.; Chandler, D.; Weeks, J. D. Hydrophobicity at Small and Large Length Scales.
 J. Phys. Chem. B 1999, 103, 4570–4577.
- (6) Huang, D. M.; Chandler, D. Temperature and Length Scale Dependence of Hydrophobic Effects and their Possible Implications for Protein Folding. *Proc. Natl. Acad. Sci. U.S.A.* 2000, *97*, 8324–8327.
- (7) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. Nature 2005, 437, 640–647.
- (8) Garde, S.; Patel, A. J. Unraveling the Hydrophobic Effect, one Molecule at a Time. Proc. Natl. Acad. Sci. U.S.A. 2011, 108, 16491–16492.
- (9) Li, I. T. S.; Walker, G. C. Signature of Hydrophobic Hydration in a Single Polymer. Proc. Natl. Acad. Sci. U.S.A. 2011, 108, 16527–16532.

- (10) Wang, J.; Bratko, D.; Luzar, A. Probing Surface Tension Additivity on Chemically Heterogeneous Surfaces by a Molecular Approach. Proc. Natl. Acad. Sci. U.S.A. 2011, 108, 6374–6379.
- (11) Mittal, J.; Hummer, G. Static and Dynamic Correlations in Water at Hydrophobic Interfaces. Proc. Natl. Acad. Sci. U.S.A. 2008, 105, 20130–20135.
- (12) Pratt, L. R.; Chandler, D. Theory of the Hydrophobic Effect. J. Chem. Phys. 1977, 67, 3683–3704.
- (13) Hummer, G.; Garde, S.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. Hydrophobic Effects on a Molecular Scale. J. Phys. Chem. B 1998, 102, 10469–10482.
- (14) Ghosh, T.; Garcia, A. E.; ; Garde, S. Molecular Dynamics Simulations of Pressure Effects on Hydrophobic Interactions. J. Am. Chem. Soc. 2001, 123, 10997–11003.
- (15) Rajamani, S.; Truskett, T. M.; Garde, S. Hydrophobic Hydration from Small to Large Lengthscales: Understanding and Manipulating the Crossover. *Proc. Natl. Acad. Sci.* U.S.A. 2005, 102, 9475–9480.
- (16) Ashbaugh, H. S.; Pratt, L. R. Colloquium: Scaled Particle Theory and the Length Scales of Hydrophobicity. *Rev. Mod. Phys.* 2006, 78, 159–178.
- (17) Sarupria, S.; Garde, S. Quantifying Water Density Fluctuations and Compressibility of Hydration Shells of Hydrophobic Solutes and Proteins. *Phys. Rev. Lett.* 2009, 103, 037803.
- (18) Patel, A. J.; Varilly, P.; Chandler, D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. J. Phys. Chem. B 2010, 114, 1632–1637.
- (19) Jamadagni, S. N.; Godawat, R.; Garde, S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. Annu. Rev. Chem. Biomol. Eng. 2011, 2, 147–171.

- (20) Patel, A. J.; Varilly, P.; Jamadagni, S. N.; Acharya, H.; Garde, S.; Chandler, D. Extended Surfaces Modulate Hydrophobic Interactions of Neighboring Solutes. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 17678–17683.
- (21) Hummer, G.; Garde, S.; Garcia, A. E.; Pohorille, A.; Pratt, L. R. An Information Theory Model of Hydrophobic Interactions. *Proc. Natl. Acad. Sci. U.S.A.* 1996, 93, 8951–8955.
- (22) Patel, A.; Varilly, P.; Chandler, D.; Garde, S. Quantifying Density Fluctuations in Volumes of All Shapes and Sizes Using Indirect Umbrella Sampling. J. Stat. Phys. 2011, 145, 265–275.
- (23) Vaikuntanathan, S.; Geissler, P. L. Putting Water on a Lattice: The Importance of Long Wavelength Density Fluctuations in Theories of Hydrophobic and Interfacial Phenomena. *Phys. Rev. Lett.* **2014**, *112*, 020603.
- (24) Jin, Z.; Kim, J.; Wu, J. Shape Effect on Nanoparticle Solvation: A Comparison of Morphometric Thermodynamics and Microscopic Theories. *Langmuir* 2012, 28, 6997– 7006.
- (25) Ashbaugh, H. S. Scaled-particle Theory Analysis of Cylindrical Cavities in Solution. *Phys. Rev. E* 2015, *91*, 042315.
- (26) Grigoriev, F. V.; Basilevsky, M. V.; Gabin, S. N.; Romanov, A. N.; Sulimov, V. B. Cavitation Free Energy for Organic Molecules Having Various Sizes and Shapes. J. Phys. Chem. B 2007, 111, 13748–13755.
- (27) Harris, R. C.; Pettitt, B. M. Effects of Geometry and Chemistry on Hydrophobic Solvation. Proc. Natl. Acad. Sci. U.S.A. 2014, 111, 14681–14686.
- (28) Alinchenko, M. G.; Anikeenko, A. V.; Medvedev, N. N.; Voloshin, V. P.; Mezei, M.;

Jedlovszky, P. Morphology of Voids in Molecular Systems. A Voronoi-Delaunay Analysis of a Simulated DMPC Membrane. J. Phys. Chem. B 2004, 108, 19056–19067.

- (29) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. J. Chem. Phys. 2004, 120, 9665.
- (30) Babin, V.; Leforestier, C.; Paesani, F. Development of a "First Principles" Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. J. Chem. Theory Comp. 2013, 9, 5395–5403.
- (31) Babin, V.; Medders, G. R.; Paesani, F. Development of a First Principles Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. J. Chem. Theory Comp. 2014, 10, 1599–1607.
- (32) Medders, G. R.; Babin, V.; Paesani, F. Development of a First-Principles Water Potential with Flexible Monomers. III. Liquid Phase Properties. J. Chem. Theory Comp. 2014, 10, 2906–2910.
- (33) Gasparotto, P.; Hassanali, A. A.; Ceriotti, M. Probing Defects and Correlations in the Hydrogen-Bond Network of ab Initio Water. J. Chem. Theory Comp. 2016, 12, 1953–1964.
- (34) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. J. Comp. Phys. 1995, 117, 1–19.
- (35) Wang, Z.-J.; Valeriani, C.; Frenkel, D. Homogeneous Bubble Nucleation Driven by Local Hot Spots: A Molecular Dynamics Study. J. Phys. Chem. B 2009, 113, 3776–3784.
- (36) Gonzlez, M. A.; Menzl, G.; Aragones, J. L.; Geiger, P.; Caupin, F.; Abascal, J. L. F.; Dellago, C.; Valeriani, C. Detecting Vapour Bubbles in Simulations of Metastable Water. J. Chem. Phys. 2014, 141, 18C511.

- (37) Structured grids differ from unstructured ones mainly in that the former are regularly connected, while the latter are charecterized by an irregular connectivity, that is the way by which the vertices of the grid specify the volume elements. See e.g. Ref. 54.
- (38) Jedlovszky, P. The Local Structure of Various Hydrogen Bonded Liquids: Voronoi Polyhedra Analysis of Water, Methanol, and HF. J. Chem. Phys. 2000, 113, 9113– 9121.
- (39) Shih, J.; Sheu, S.; Mou, C. A Voronoi Polyhedra Analysis of Structures of Liquid Water.
 J. Chem. Phys. 1994, 100, 2202–2212.
- (40) Medvedev, N. N.; Voloshin, V. P.; Luchnikov, V. A.; Gavrilova, M. L. An Algorithm for Three-dimensional Voronoi S-network. J. Comp. Chem. 2006, 27, 1676–1692.
- (41) The VNP code. http://www.kinetics.nsc.ru/mds/?Software:VNP, Accessed: 05-07-2016.
- (42) Voloshin, V. P.; Medvedev, N. N.; Andrews, M. N.; Burri, R. R.; Winter, A., Roland Geiger Volumetric Properties of Hydrated Peptides: Voronoi-Delaunay Analysis of Molecular Simulation Runs. J. Phys. Chem. B 2011, 115, 14217–14228.
- (43) Voloshin, V. P.; Kim, A. V.; Medvedev, N. N.; Winter, A., Roland Geiger Calculation of the Volumetric Characteristics of Biomacromolecules in Solution by the Voronoi-Delaunay technique. *Biophys. Chem.* **2014**, *192*, 1–9.
- (44) Sosso, G. C.; Miceli, G.; Caravati, S.; Giberti, F.; Behler, J.; Bernasconi, M. Fast Crystallization of the Phase Change Compound GeTe by Large-Scale Molecular Dynamics Simulations. J. Phys. Chem. Lett. 2013, 4, 4241–4246.
- (45) Hassanali, A. A.; Giberti, F.; Sosso, G. C.; Parrinello, M. The Role of the Umbrella Inversion Mode in Proton Diffusion. *Chem. Phys. Lett.* 2014, 599, 133–138.

- (46) An alpha shape associated with a set of points is the generalization of the concept of convex hull (see e.g. Ref. 55).
- (47) MATLAB, version 7.10.0 (R2010a); The MathWorks Inc.: Natick, Massachusetts, 2010.
- (48) Ruocco, G.; Sampoli, M.; Vallauri, R. Analysis of the Network Topology in Liquid Water and Hydrogen Sulphide by Computer Simulation. J. Chem. Phys. 1992, 96, 6167–6176.
- (49) Godawat, R.; Jamadagni, S. N.; Garde, S. Characterizing hydrophobicity of Interfaces by Using Cavity Formation, Solute Binding, and Water Correlations. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106, 15119–15124.
- (50) Garde, S.; Hummer, G.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. Origin of Entropy Convergence in Hydrophobic Hydration and Protein Folding. *Phys. Rev. Lett.* 1996, 77, 4966.
- (51) Vaikuntanathan, S.; Rotskoff, G.; Hudson, A.; Geissler, P. L. Necessity of Capillary Modes in a Minimal Model of Nanoscale Hydrophobic Solvation. *Proc. Natl. Acad. Sci.* U.S.A. 2016, 113, E2224–E2230.
- (52) Hassanali, A.; Giberti, F.; Cuny, J.; Khne, T. D.; Parrinello, M. Proton Transfer Through the Water Gossamer. Proc. Natl. Acad. Sci. U.S.A. 2013, 110, 13723–13728.
- (53) Gibb, C. L. D.; Gibb, B. C. Anion Binding to Hydrophobic Concavity Is Central to the Salting-in Effects of Hofmeister Chaotropes. J. Am. Chem. Soc. 2011, 133, 7344–7347.
- (54) Thompson, J.; Soni, B.; Weatherill, N. Handbook of grid generation; Boca Raton; London: CRC Press, 1999.
- (55) Skiena, S. The algorithm design manual; Santa Clara, Calif: TELOS The Electronic Library of Science, 1998.

Graphical TOC Entry



On the Role of Non-Spherical Cavities in Short Length-Scale Density Fluctuations in Water SUPPLEMENTARY INFORMATION

Gabriele Cesare Sosso,^{*,†} Sebastiano Caravati,[‡] Grant Rotskoff,[¶] Suriyanarayan Vaikuntanathan,[§] and Ali Hassanali^{*,||}

†Thomas Young Centre, London Centre for Nanotechnology and Department of Physics and Astronomy, University College London, Gower Street London WC1E 6BT, United Kingdom ‡Department of Chemistry, University of Zurich, Winterhurerstrasse 190 CH-8057, Switzerland

¶Biophysics Graduate Group, University of California, Berkeley, CA 94720, USA §Department of Chemistry, The University of Chicago, IL 60637, USA

||Condensed Matter and Statistical Physics Section, The Abdus Salam International Centre for Theoretical Physics, I-34151 Trieste, Italy

E-mail: g.sosso@ucl.ac.uk; ahassana@ictp.it

We provide supplementary information about:

- The probability P(n) of finding n solvent centers within spherical cavities (SC) and Voronoi-Delaunay Voids (VDV)
- The lattice model results with respect to SC and VDV
- The scaling exponent of the excess free energy of solvation as a function of volume and surface area
- The structural properties of the water network in the neighborhood of the VDV
- The statistics of VDV as obtained for the solvation of alanine in water
- VDV and SC in MB-pol water

P(n) for SC and VDV: Raw Data and Gaussian Fits

In Fig. S1 and Fig. S2 we show the probability P(n) of observing *n* solvent centers (oxygens atoms in this case) within a SC or a VDV of a given volume. We present the data obtained by using configurations coming either from ab initio or classical molecular dynamics (see main text), together with the corresponding Gaussian fittings. The results refer to the selected voids discussed in the main text (Fig. 5). While small differences in the P(n) exist between the ab initio (128-molecule, $\sim 10^1$ ps) and the classical (4096-molecule, $\sim 10^4$ ps) data, VDV consistently display softer tails with respect to the correspondent SC.

The Lattice Model

As discussed in the main text, including lattice fluctuations^{1,2} on top of Gaussian contributions when dealing with density fluctuations of small length scales results in substantial errors. This is shown in Fig. S3, where we compare the P(n) we have directly obtained from



Fig. S1: Probability P(n) of observing n solvent centers within a SC or a VDV of a given volume, here represented by the radius (r, purple labels) of a sphere of the same volume. Left and right panels refer to the results obtained by ab initio and classical molecular dynamics simulations respectively. The discrete values of P(n) (points) have been fitted by Gaussian distributions (lines), also reported.

our molecular dynamics simulations (VDV/SC_{T4}) with those calculated taking into account both the Gaussian and the lattice contribution (VDV/SC,GL). While the qualitative trend of VDV displaying softer P(n) tails compared to SC still holds, significant discrepancies between VDV/SC_{T4} and VDV/SC,GL can be observed.



Fig. S2: Probability P(n) of observing n solvent centers within a SC or a VDV of a given volume, here represented by the radius (r, purple labels) of a sphere of the same volume. Left and right panels refer to the results obtained by ab initio and classical molecular dynamics simulations respectively. The discrete values of P(n) (points) have been fitted by Gaussian distributions (lines), also reported.

Gaussian Statistics: Scaling Exponents

In the hypothesis that Gaussian statistics is sufficient to describe both SC and VDV, the following expression for the excess free energy₄ $\Delta \mu$ holds:³



Fig. S3: Probability P(n) of observing n solvent centers (n oxygens) within VDV (VDV_{T4}) and a corresponding SC (SC_{T4}) of the same volume. Statistics has been accumulated over 10^5 configurations of a 4096-mol model of TIP4P/Ew water. The results obtained within the Gaussian approximation plus the lattice contribution (see text) for VDV (VDV,GL) and SC (SC,GL) are also reported as dotted lines.

$$\Delta \mu \sim k_B T \cdot \frac{\rho_w^2 V^2}{2\chi_V} + k_B T \cdot \frac{\ln(2\pi\chi_V)}{2} \tag{1}$$

where ρ_w , V and χ_V are the water density, the volume of the VDV/SC and the variance of P(n) for that particular VDV/SC respectively. The second term of Eq. 1, which depends weakly (logarithmically) on χ_V , is more often than not neglected. In addition, for SC $\chi_V \sim V$,⁴ so that $\Delta \mu_{SC} \sim \frac{k_B T}{2} \cdot \rho_w^2 V$. The latter expression has been used to obtain the data fit (SC_{T4,nolog}) reported in Fig. S4a. Note that including the second term of Eq. 1, so that $\Delta \mu_{SC} \sim \frac{k_B T}{2} \cdot [\rho_w^2 V + \ln(2\pi V)]$, results in a data fit (SC_{T4,log} in Fig. S4a) in very poor agreement with our values of $\Delta \mu$. As for VDV, the scaling exponent for χ_V is unknown. Thus, we have fitted the excess free energy of the VDV using the following expression:

$$\Delta \mu_{VDV} \sim k_B T \cdot \frac{\rho_w^2 V^2}{2V^\alpha} \tag{2}$$

The result is reported in Fig. S4a. We have obtained $\alpha = 1.07 \pm 0.01$, slightly larger than $\alpha_{SC}=1$. At this point, one can also compute directly χ_V as the variance of the Gaussian distributions by which we have fitted the P(n) for SC and VDV with (see Fig. S1 and Fig. S2). As shown in Fig. S4b, χ_V for SC and VDV are in very good agreement with $\chi_V^{SC} = V$ and



Fig. S4: Scaling exponents. The same SC and VDV considered in the main text (see e.g. Fig. 5) have been considered. Statistics has been accumulated over 10^5 configurations of a 4096-mol model of TIP4P/Ew water. a) Excess free energy $\Delta\mu$ as a function of the volume for SC (diamonds) and VDV (triangles). The data fits obtained by means of Eq. 1 with (SC/VDV_{T4,log}) or without (SC/VDV_{T4,nolog}) the logarithmic contribution (see text) are also reported.. b) Variance of P(n) (χ_V) , obtained from the Gaussian fit of P(n) for SC (diamonds) and VDV (triangles), as a function of volume. The scaling trends $\chi_V^{SC} \sim V$ (dotted, light blue line) and $\chi_V^{VDV} \sim V^{1.07}$ (solid, magenta line) are also reported. c) Surface area S_A as a function of volume for SC and VDV. The scaling trends $S_A^{SC} \sim V^{2/3}$ (solid line) and $S_A^{VDV} \sim V^{0.82}$ (dotted line) are also reported.

 $\chi_V^{VDV} = V^{1.07}$ respectively. We also report in Fig. S4c the scaling trend of the surface area S_A as a function of the volume of SC and VDV. Note the much stronger dependence for VDV if compared to SC. Finally, in Fig. S5 we illustrate the volume dependence of $\Delta \mu / S_A$ for VDV vs SC, including the scaling exponents as obtained from the above expressions for $\Delta \mu (V)$ and $S_A(V)$.

Structural Properties of the Water Network around VDV

In Fig. S6 we report the in-degree (the number of HB being accepted by a water molecule) and the out-degree (the number of HB being donated) of water molecules that are within 2 Å of the vertices of the Voronoi-polyhedra enclosing the VDV. Looking at the distributions of the in and out degrees suggest that most of these waters are tetrahedral - there is no signature of an increase in the number of defect water molecules such as over or under-coordinated water molecules.



Fig. S5: Scaling exponents. The same SC and VDV considered in the main text (see e.g. Fig. 5) have been considered. Statistics has been accumulated over 10^5 configurations of a 4096-mol model of TIP4P/Ew water. a) $\Delta \mu/S_A$ as a function of volume for VDV vs SC. The volume dependencies, as obtained from the scaling exponents discussed in the text and depicted in Fig S4, are reported in panel b).



Fig. S6: Probability density distribution of a) in-degree and b) out-degree water molecules that are within 2 Å of the vertices of the Voronoi-polyhedra enclosing the VDV.

VDV Statistics for the Solvation of Alanine

In Fig. S7 we show the probability density distribution functions of volume, surface area and asphericity for the VDV created by removing the alanine molecule from the water network.



Fig. S7: Probability density distribution functions of a) volume b) surface area and c) asphericity of the VDV created by removing the alanine molecule from the water network. Results are averaged over $\sim 10^3$ configurations of a 128-mol model of BLYP water.



Fig. S8: a) Cumulative distribution function (CDF) of the volume of VDV and SC within the 216mol model of MB-pol water at room temperature and pressure. b) Surface area S_A as a function of volume for SC and VDV. c) Asphericity of VDV as a function of volume. Reference values for different regular shapes are also reported: SC, TH, CUB and L_{CUB} refer to sphere, tetrahedron, cube and asymmetrical cuboid. d) Excess chemical potential $\Delta\mu$ as a function the radius r of SC and VDV - as obtained from the MB-pol trajectory. For VDV, r is defined as the radius of a sphere with exactly the same volume as the VDV. Note however that $\Delta\mu$ has been obtained from the P(n) calculated using actual VDV with shapes like those depicted in Fig. 1 and Fig. 3 - main text.

VDV and SC in MB-pol water

We report in Fig. S8 some additional information about VDV and SC as obtained within a 80 ps long molecular dynamics trajectory at ambient temperature and pressure employing the MB-pol model of water.^{5–7} 216 water molecules have been used, sampling the NVT ensemble. We do not report the computational details in full as Fig. S8 is included with the only intent to provide further evidence that our findings do not depend on the particular water model chosen.

References

- Vaikuntanathan, S.; Geissler, P. L. Putting Water on a Lattice: The Importance of Long Wavelength Density Fluctuations in Theories of Hydrophobic and Interfacial Phenomena. *Phys. Rev. Lett.* **2014**, *112*, 020603.
- (2) Vaikuntanathan, S.; Rotskoff, G.; Hudson, A.; Geissler, P. L. Necessity of Capillary Modes in a Minimal Model of Nanoscale Hydrophobic Solvation. *Proc. Natl. Acad. Sci.* U.S.A. 2016, 113, E2224–E2230.
- (3) Garde, S.; Hummer, G.; Garcia, A. E.; Paulaitis, M. E.; Pratt, L. R. Origin of Entropy Convergence in Hydrophobic Hydration and Protein Folding. *Phys. Rev. Lett.* **1996**, 77, 4966.
- (4) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. Nature 2005, 437, 640–647.
- (5) Babin, V.; Leforestier, C.; Paesani, F. Development of a First Principles Water Potential with Flexible Monomers: Dimer Potential Energy Surface, VRT Spectrum, and Second Virial Coefficient. J. Chem. Theory Comput. 2013, 9, 5395–5403.
- (6) Babin, V.; Medders, G. R.; Paesani, F. Development of a First Principles Water Potential with Flexible Monomers. II: Trimer Potential Energy Surface, Third Virial Coefficient, and Small Clusters. J. Chem. Theory Comput. 2014, 10, 1599–1607.
- Medders, G. R.; Babin, V.; Paesani, F. Development of a First-Principles Water Potential with Flexible Monomers. III. Liquid Phase Properties. J. Chem. Theory Comput. 2014, 10, 2906–2910.