

RESEARCH ARTICLE | MAY 22 2024

## Local number fluctuations in ordered and disordered phases of water across temperatures: Higher-order moments and degrees of tetrahedrality

Michael A. Klatt  ; Jaeuk Kim  ; Thomas E. Gartner, III  ; Salvatore Torquato  

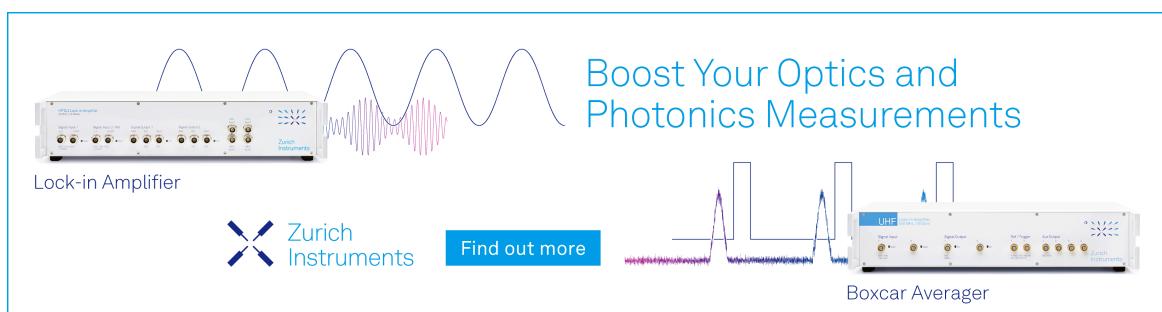
 Check for updates

*J. Chem. Phys.* 160, 204502 (2024)

<https://doi.org/10.1063/5.0204696>

  
View  
Online

  
Export  
Citation



Boost Your Optics and Photonics Measurements

Lock-in Amplifier

Zurich Instruments

Find out more

Boxcar Averager

The advertisement features a Zurich Instruments UHF Lock-in Amplifier and a Boxcar Averager. The Lock-in Amplifier is shown with a sine wave input and a multi-frequency sine wave output. The Boxcar Averager is shown with a signal input and a digital output. The text 'Boost Your Optics and Photonics Measurements' is displayed in blue. The Zurich Instruments logo is present, and a 'Find out more' button is visible.

# Local number fluctuations in ordered and disordered phases of water across temperatures: Higher-order moments and degrees of tetrahedrality

Cite as: *J. Chem. Phys.* **160**, 204502 (2024); doi: [10.1063/5.0204696](https://doi.org/10.1063/5.0204696)

Submitted: 22 February 2024 • Accepted: 30 April 2024 •

Published Online: 22 May 2024



View Online



Export Citation



CrossMark

Michael A. Klatt,<sup>1,2,3,a</sup> Jaeuk Kim,<sup>4,5,6,b</sup> Thomas E. Gartner III,<sup>7,c</sup> and Salvatore Torquato<sup>4,5,6,8,d</sup>

## AFFILIATIONS

<sup>1</sup> German Aerospace Center (DLR), Institute for AI Safety and Security, Wilhelm-Runge-Str. 10, 89081 Ulm, Germany

<sup>2</sup> German Aerospace Center (DLR), Institute for Material Physics in Space, 51170 Köln, Germany

<sup>3</sup> Department of Physics, Ludwig-Maximilians-Universität München, Schellingstr. 4, 80799 Munich, Germany

<sup>4</sup> Department of Chemistry, Princeton University, Princeton, New Jersey 08544, USA

<sup>5</sup> Department of Physics, Princeton University, Princeton, New Jersey 08544, USA

<sup>6</sup> Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544, USA

<sup>7</sup> Department of Chemical and Biomolecular Engineering, Lehigh University, Bethlehem, Pennsylvania 18015, USA

<sup>8</sup> Program in Applied and Computational Mathematics, Princeton University, Princeton, New Jersey 08544, USA

**Note:** This paper is part of the JCP Special Topic on Water: Molecular Origins of its Anomalies.

<sup>a</sup>[michael.klatt@dlr.de](mailto:michael.klatt@dlr.de)

<sup>b</sup>[jaeukk@princeton.edu](mailto:jaeukk@princeton.edu)

<sup>c</sup>[teg323@lehigh.edu](mailto:teg323@lehigh.edu)

<sup>d</sup>Author to whom correspondence should be addressed: [torquato@princeton.edu](mailto:torquato@princeton.edu)

## ABSTRACT

The isothermal compressibility (i.e., related to the asymptotic number variance) of equilibrium liquid water as a function of temperature is minimal under near-ambient conditions. This anomalous non-monotonic temperature dependence is due to a balance between thermal fluctuations and the formation of tetrahedral hydrogen-bond networks. Since tetrahedrality is a many-body property, it will also influence the higher-order moments of density fluctuations, including the skewness and kurtosis. To gain a more complete picture, we examine these higher-order moments that encapsulate many-body correlations using a recently developed, advanced platform for local density fluctuations. We study an extensive set of simulated phases of water across a range of temperatures (80–1600 K) with various degrees of tetrahedrality, including ice phases, equilibrium liquid water, supercritical water, and disordered nonequilibrium quenches. We find clear signatures of tetrahedrality in the higher-order moments, including the skewness and excess kurtosis, which scale for all cases with the degree of tetrahedrality. More importantly, this scaling behavior leads to non-monotonic temperature dependencies in the higher-order moments for both equilibrium and non-equilibrium phases. Specifically, under near-ambient conditions, the higher-order moments vanish most rapidly for large length scales, and the distribution quickly converges to a Gaussian in our metric. However, under non-ambient conditions, higher-order moments vanish more slowly and hence become more relevant, especially for improving information-theoretic approximations of hydrophobic solubility. The temperature non-monotonicity that we observe in the full distribution across length scales could shed light on water's nested anomalies, i.e., reveal new links between structural, dynamic, and thermodynamic anomalies.

Published under an exclusive license by AIP Publishing. <https://doi.org/10.1063/5.0204696>

## I. INTRODUCTION

Physical properties of many-particle systems can be greatly influenced by their density fluctuations.<sup>1–9</sup> The relationship between the isothermal compressibility and infinite-wavelength density fluctuations in thermal equilibrium is a classic example.<sup>3,10</sup> More generally, in any many-particle system, both in and out of equilibrium, local density fluctuations can be comprehensively quantified via the probability distribution  $P[N(R)]$ , where  $N(R)$  is the number of particles in a spherical observation window of radius  $R$ ,<sup>11,12</sup> as shown in Fig. 1. The variance of this probability distribution, i.e., the number variance  $\sigma_N^2(R)$ , and its large- $R$  scaling are key determinants of the physical and structural properties of many-body systems, including elastic moduli, electronic transport properties, and electromagnetic properties.<sup>1,3,10,13–15</sup> Controlling  $\sigma_N^2(R)$  can lead to optimal mechanical responses,<sup>16</sup> optimal transport properties,<sup>17–19</sup> and superior strategies for sensing<sup>20</sup> or learning.<sup>21</sup>

The fact that water is so abundant and of critical importance in biological and industrial contexts has motivated many investigators to understand and quantify the relationships between its complex physical properties and structure, especially its density fluctuations.<sup>22–32</sup> Water's intricate local structure exhibits temperature- and pressure-dependent shifts in a generally tetrahedral hydrogen bond network, which strongly affects the density fluctuations of water. A balance between thermal fluctuations and the formation of such a tetrahedral network is believed to be the reason for liquid water to anomalously exhibit extrema in thermodynamic response functions upon cooling at constant pressure.<sup>33</sup> For example, the isothermal compressibility of liquid water in equilibrium as a function of temperature is minimal under near-ambient conditions, which directly translates into a corresponding minimum of the large- $R$  asymptotic number variance. Another example where the analysis of density fluctuations,

specifically  $\sigma_N^2(R)$ , has proven useful to predict the physical properties of water (mediated by its tetrahedral nature) is the solubility of hydrophobic solutes in water.<sup>34–42</sup> Finally, two-body information and tetrahedrality have also been used to study density fluctuations in amorphous ices and transitions between their different forms.<sup>43–46</sup>

Tetrahedrality is inherently a many-body property, involving at least one molecule and its four neighbors. Therefore, it affects not only second-order properties, such as  $\sigma_N^2(R)$ , but also higher-order moments of  $P[N(R)]$ , specifically, its skewness and kurtosis. To better understand the complex and anomalous behaviors in the physical properties of water, there is much to be gained via a complete description of the dependence of water's density fluctuations on its tetrahedrality. While there have been previous efforts to understand such effects by using typical two-body statistics<sup>47,48</sup> (also see Fig. 8 in Appendix A), these approaches may provide an incomplete picture because they lack the effects of tetrahedrality on three- and four-body properties. Thus, it is desired to extract information from the higher-order moments and full distribution  $P[N(R)]$  on local length scales, i.e., to go beyond the two-body statistics contained in  $\sigma_N^2(R)$ .

Here, we study in detail the higher-order moments of density fluctuations in water using a recently developed, advanced “platform” or “toolset” for local density fluctuations.<sup>11</sup> Thus, we gain a refined physical and structural understanding of the impact of tetrahedrality. This approach highlights the fundamental importance of higher-order structural information to fully characterize density fluctuations across length scales. The analysis is applicable to generic many-particle systems and is based on explicit closed-form integral expressions for structural information up to three- and four-body correlations, rigorous bounds, and high-precision numerical techniques.

We apply this new toolset to an extensive set of water states across the phase diagram with various degrees of tetrahedrality, simulated via the TIP4P/2005 model.<sup>49</sup> For this reason, we also consider water states far from ambient conditions, which have often been neglected in previous studies about water's density fluctuations. Specifically, we consider (i) liquid water in equilibrium at ambient temperature (300 K), close to the liquid–vapor critical point (646.4 K), and in a supercritical state (1600 K); see Fig. 1 for a snapshot of water molecules at  $T = 300$  K; (ii) supercooled liquid water at  $T = 200$  K; (iii) hexagonal ice Ih and cubic ice Ic just below this model's melting temperature at  $T = 250$  K; and (iv) nonequilibrium quenches of water at  $T = 80, 180, 190$ , and 200 K. Such diverse phases of water, some far from ambient, are relevant for applications such as biopreservation,<sup>50</sup> climate modeling,<sup>51,52</sup> astrobiology/space exploration,<sup>53</sup> and understanding life near hydrothermal vents.<sup>54</sup> We compare our states of water to two further reference systems with low or high degrees of tetrahedral order, respectively: equilibrium hard spheres (as a prototypical example of a simple liquid)<sup>55,56</sup> and a continuous random network (as a classical model of amorphous silicon).<sup>57</sup>

Our key findings are briefly summarized as follows:

- We show that higher-order moments contain many-body information and hence detect tetrahedrality, i.e., they exhibit distinct features that scale with the degree of tetrahedrality;

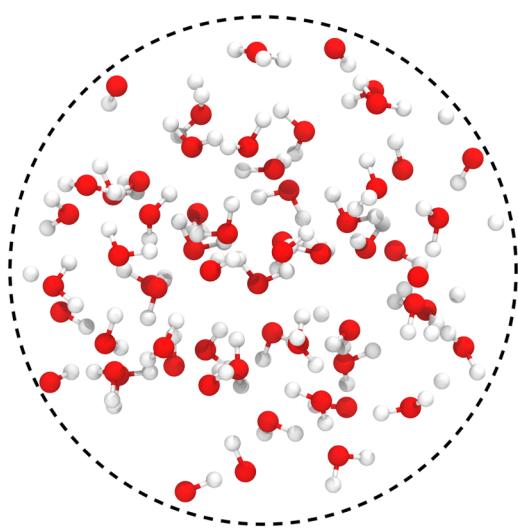


FIG. 1. Snapshot of water molecules in the TIP4P/2005 model in the equilibrium liquid at  $T = 300$  K. Density fluctuations are evaluated by the number distribution of oxygen atoms within a spherical observation window, here with radius  $R = 7$  Å.

- While lower-order moment approximations work best close to room temperatures, we demonstrate that higher-order moments are needed at both higher and lower temperatures;
- We show that higher-order moments improve information-theoretic approximations of the hydrophobic solubility at far-from-ambient conditions.

More specifically, our results reveal how the degree of tetrahedrality is captured by explicit features in the skewness and excess kurtosis at specific radii  $R$ , as shown in Sec. III A and Figs. 2–5. These features scale for all of our phases with the degree of tetrahedrality and thus offer an additional, statistically robust characterization of tetrahedral order in disordered phases of water both in and out of equilibrium.

More importantly, in Sec. III B, we find that the tetrahedrality also leads to non-monotonic temperature dependencies in the higher-order moments and, consequently, a Gaussian distance metric. While the well-known minimum in the isothermal compressibility as a function of temperature only predicts a minimal *asymptotic* variance, here we observe a non-monotonic behavior also in the higher-order moments of local density fluctuations. For intermediate to large radii, the distance to a Gaussian distribution vanishes most rapidly under near-ambient conditions. At both higher and lower temperatures, the decay slows down, as shown in Fig. 6. Hence, higher-order moments are more relevant under non-ambient than under ambient conditions. As discussed below, we surmise that this non-monotonicity again results from a balance between tetrahedrality and thermal fluctuations, but in our case, this trade-off affects the higher-order moments. The observed tendency of slower convergence cannot be predicted solely from the number variance or other two-body information.<sup>11</sup>

Consequently, the accuracy of any second-order approximation that assumes Gaussian density fluctuations will deteriorate as the temperature moves away from ambient conditions. As a prominent example, we study this effect for the information-theoretic approximations of hydrophobic solubility;<sup>23</sup> see Sec. III C. Previous studies conducted under ambient conditions found that higher-order moments do not improve upon the second-order approximation.<sup>23,58</sup> While we confirm this finding for equilibrium water at  $T = 300$  K, we show that for all other state points we consider, higher-order moments indeed improve the approximation, i.e., equilibrium water around room temperature is again an exception. More specifically, we find that the approximation is accurate only for a range of radii where the number of constrained moments essentially determines the entire probability distribution; the higher the number of moments, the larger this range of radii, as shown in Fig. 7.

The structural differences revealed by our analysis of the higher-order moments directly translate into different solvation behaviors. For example, we show that for strongly negative values of the skewness, the information-theoretic prediction of the hydrophobic solubility tends to overestimate the excess chemical potential. For a review of the far-reaching chemical implications of hydrophobic solubility, see the study by Rego and Patel.<sup>42</sup> Moreover, our structural characterization of water at all length scales and the non-monotonic temperature dependence that we discovered in the higher-order moments may reveal, in a future work, links between

the known structural, dynamic, and thermodynamic anomalies of water,<sup>59</sup> as briefly discussed in Sec. IV.

Even though experimentally accessible pair correlations might look similar for two liquid-like states of water, there can be distinct structural differences between these states with physical consequences as discussed above.<sup>59</sup> Thus, our study further motivates the need for experimental methods to ascertain three- and higher-body correlations in water systems.<sup>60–62</sup>

We provide details on the models, phases, and structural characteristics (together with their mathematical definitions) in Sec. II, before we present our results in Sec. III (as described above). Finally, we provide concluding remarks in Sec. IV.

## II. METHODS

We use extensive, state-of-the-art simulations to represent a broad range of water phases and reference states. Then, we analyze these samples with the methods developed by Torquato, Kim, and Klatt.<sup>11</sup>

### A. Models and phases

We performed isothermal–isochoric (NVT) molecular dynamics (MD) simulations of water using the TIP4P/2005 potential, one of the most widely used water models.<sup>49</sup> TIP4P/2005 is a rigid 4-site classical water model that faithfully reproduces water's properties and phase diagram across a broad range of states.<sup>63</sup> We used GROMACS v2018.4,<sup>64</sup> integrated the equations of motion with a leap-frog algorithm with time step 2 fs, and used a stochastic velocity-rescaling thermostat with relaxation time 0.1 ps for temperature control. We enforced bond and angle constraints with a sixth-order LINear Constraint Solver (LINCS) algorithm, and the van der Waals and real-space Coulomb cutoff distances were 1.2 nm. We used a particle-mesh Ewald scheme to treat long-range electrostatics with a Fourier grid spacing of 0.16 nm. We prepared disordered initial configurations using the gmx solvate method at the given simulation density.

For liquid water at  $T = 300$  K, we equilibrated the system for 1 ns and collected frames for analysis every subsequent 100 ps. For  $T = 200$  K metastable supercooled liquid water, we equilibrated the system for 400 ns and collected frames for analysis every subsequent 200 ps. For both liquid water systems, the mass density was 1.0 g/cm<sup>3</sup>. We note that while at  $T = 200$  K, liquid water is metastable to ice I due to the separation of timescales between structural equilibration and ice nucleation in finite-size simulations, it is possible to prepare structurally equilibrated liquid water even at deep supercoolings. For supercritical water at  $T = 646.4$  K (1.01 times the liquid–vapor critical temperature of 640 K<sup>65</sup>), we equilibrated the system for 100 ns and collected frames for analysis every subsequent 100 ps. For supercritical water at  $T = 1600$  K, we used a smaller time step size of 1 fs, equilibrated the system for 10 ns, and collected frames for analysis every subsequent 100 ps. The supercritical systems were performed at the critical density of 0.31 g/cm<sup>3</sup>.<sup>65</sup> For ice Ic and Ih, we prepared proton-disordered initial configurations using the GenIce package<sup>66</sup> at densities of 0.944 g/cm<sup>3</sup> for ice Ic and 0.921 g/cm<sup>3</sup> for ice Ih.<sup>49,67</sup> We equilibrated the ice systems for 100 ps and collected frames for analysis every subsequent 100 ps. The ice simulations were performed at  $T = 250$  K, just below their melting

**TABLE I.** Simulation parameters for all water states of the TIP4P/2005 model considered here, where  $T$  is the temperature;  $N$  is the number of oxygen atoms inside the simulation box;  $N_c$  is the number of configurations;  $L_x$ ,  $L_y$ , and  $L_z$  are the side lengths of the simulation box (where  $*$  indicates a cubic simulation box); and  $\rho$  is the number density.

Phases		$T$ (K)	$N$	$N_c$	$L_x$ (nm)	$L_y$ (nm)	$L_z$ (nm)	$\rho$ (nm $^{-3}$ )	$\rho^{1/3}$ (nm)
Solid	Ice Ih	250.0	35 152	500	10.17	9.57	11.74	30.8	0.319
	Ice Ic	250.0	32 768	500	10.21	*	*	30.8	0.319
Equilibrium liquid		200.0	36 424	500	10.29	*	*	33.4	0.311
		300.0	36 424	500	10.29	*	*	33.4	0.311
Quench		80.0	36 424	500	10.29	*	*	33.4	0.311
		180.0	36 424	500	10.29	*	*	33.4	0.311
		190.0	36 424	500	10.29	*	*	33.4	0.311
		200.0	36 424	500	10.29	*	*	33.4	0.311
Supercritical fluid		646.4	36 424	500	15.20	*	*	10.4	0.458
		1600.0	36 424	500	15.20	*	*	10.4	0.458

temperature of  $T \approx 252$  K.<sup>49</sup> For the quench configurations, we followed exactly the stepwise quench procedure given in the study by Gartner *et al.*<sup>45</sup> at a cooling rate of 10 K/ns, except in this work, we performed the quenches at constant volume corresponding to a mass density of 1.0 g/cm<sup>3</sup> instead of constant pressure. Each configuration for analysis was taken from an independent quench simulation. The system sizes for all simulations are presented in Table I.

To compare liquid water at  $T = 300$  K to a simple liquid, we simulate an equilibrium hard-sphere liquid with a packing fraction  $\phi = 31.7\%$  and a particle number  $N = 10^5$  via the Monte Carlo method.<sup>68</sup> We chose the value of  $\phi$  to correspond to an effective packing fraction of liquid water under near-ambient conditions<sup>59,70</sup> in the following sense. At unit number density, the hard-sphere diameter  $D$  is equal to the smallest distance at which the pair correlation function of two oxygen atoms reaches one, i.e.,  $g_2(r = D) = 1$ ; for a definition of  $g_2(r)$ , see Sec. II B 2. To compare the ice phases to reference systems with high tetrahedrality, we consider amorphous silicon represented by a continuous random network with 100 000 vertices from Barkema and Mousseau.<sup>57</sup>

## B. Structural characterization

We quantify the density fluctuation via  $P[N(R)]$  that incorporates many-body correlations  $g_n$  at unit number density. This structural analysis is complemented by a tetrahedral order parameter  $q$ .

### 1. Probability distribution $P[N(R)]$

We quantify density fluctuations of many-particle systems by measuring the probability distribution function  $P[N(R)]$  via the Monte Carlo method adopted in the study by Torquato, Kim, and Klatt.<sup>11</sup> Specifically, at each window radius  $R$ , we measure the values of  $N(R)$  from  $N_w(R)$  randomly placed windows in every sample configuration under periodic boundary conditions. To reduce systematic errors due to oversampling, we choose  $N_w(R)$  such that the union volume of the observation windows of radius  $R$  cannot exceed 50% of the volume  $V$  of the simulation box, i.e.,  $1 - \exp[-N_w(R)v_1(R)/V] < 0.5$ , as in the study by Torquato, Kim, and Klatt,<sup>11</sup> but with the slight improvement that we here optimize

$N_w(R)$  for each radius separately. From the determined  $P[N(R)]$ , we then estimate the first four central moments associated with number variance  $\sigma_N^2(R)$ , skewness  $\gamma_1(R)$ , and excess kurtosis  $\gamma_2(R)$ , defined, respectively, as

$$\sigma_N^2(R) := \langle [N(R) - \langle N(R) \rangle]^2 \rangle, \quad (1)$$

$$\gamma_1(R) := \langle [N(R) - \langle N(R) \rangle]^3 \rangle / [\sigma_N^2(R)]^{3/2}, \quad (2)$$

$$\gamma_2(R) := \langle [N(R) - \langle N(R) \rangle]^4 \rangle / [\sigma_N^2(R)]^2 - 3, \quad (3)$$

where  $\langle \cdot \rangle$  represents an ensemble average. The skewness  $\gamma_1(R) \in (-\infty, \infty)$  measures the asymmetry of the probability distribution  $P[N(R)]$ , i.e., a positive value of the skewness  $\gamma_1(R)$  means that the right tail is heavier than the left tail, and a negative  $\gamma_1(R)$  implies a heavier left than right tail. A zero value of  $\gamma_1(R)$  is consistent with  $P[N(R)]$  being symmetric around the mean value  $\langle N(R) \rangle$ . The excess kurtosis  $\gamma_2(R) \in [-3, \infty)$  measures how heavy the tails of  $P[N(R)]$  are compared to a Gaussian distribution. Specifically, a positive value of the excess kurtosis  $\gamma_2(R)$  means that the tails are heavier than Gaussian, and a negative value implies that the tails are lighter. For a Gaussian distribution, both the skewness and excess kurtosis are identically zero. To measure the deviations of  $P[N(R)]$  from a Gaussian distribution, we also compute the  $l_2$  distance metric,<sup>11</sup> defined as

$$l_2(R) := \left\{ [\sigma_N^2(R)]^{-1/2} \sum_{n=0}^{\infty} |F_G(n) - F_N(n)|^2 \right\}^{1/2}, \quad (4)$$

where  $F_N(n)$  and  $F_G(n)$  are the cumulative distribution functions of  $P[N(R)]$  and the discrete Gaussian distribution with mean value  $\langle N(R) \rangle$  and variance  $\sigma_N^2(R)$ , respectively.

### 2. Many-body correlation functions $g_n$

Another route to characterize density fluctuations of many-particle systems in  $\mathbb{R}^d$  is to use the  $n$ -body correlation functions

$g_n(\mathbf{r}^n)$  [or  $n$ -particle probability density functions  $\rho_n(\mathbf{r}^n)$ ] for  $n \geq 2$ , where  $\mathbf{r}^n$  is a shorthand notation<sup>3,68</sup> for the position vectors of any  $n$  points, i.e.,  $\mathbf{r}^n := \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ . The quantity  $\rho_n(\mathbf{r}^n)d\mathbf{r}^n$  is proportional to the probability of finding *any*  $n$  particles with configuration  $\mathbf{r}^n$  in volume element  $d\mathbf{r}^n := d\mathbf{r}_1 d\mathbf{r}_2 \dots d\mathbf{r}_n$  (see the book by Chiu *et al.*<sup>71</sup> for a mathematical definition). When the systems are statistically homogeneous,  $\rho_n(\mathbf{r}^n)$  is translationally invariant and hence depends only on the relative displacements, say with respect to  $\mathbf{r}_1$ :  $\rho_n(\mathbf{r}^n) = \rho_n(\mathbf{r}_{12}, \mathbf{r}_{13}, \dots, \mathbf{r}_{1n})$ , where  $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ . In particular, the one-particle function is identical to the constant *number density*  $\rho$ , and thus, it is convenient to define the  $n$ -body correlation function,

$$g_n(\mathbf{r}^n) = \rho_n(\mathbf{r}^n)/\rho^n. \quad (5)$$

In systems without a long-range order,  $g_n(\mathbf{r}^n) \rightarrow 1$  when the points  $\mathbf{r}^n$  are mutually far from one another. Thus, the deviation of  $g_n$  from unity measures the degree of spatial correlation between the particles. As in Torquato, Kim, and Klatt,<sup>11</sup> here we compare all phases at unit number density or, equivalently, rescale them by a length scale  $\rho^{-1/3}$ , which corresponds for most state points to about 3 Å (except for the supercritical fluids, where it is about 5 Å); see Table I.

The important two-body function  $g_2(\mathbf{r}_{12})$  is usually called the *pair correlation function*. The *total correlation function*  $h(\mathbf{r}_{12})$  is defined as  $h(\mathbf{r}_{12}) = g_2(\mathbf{r}_{12}) - 1$ , and thus is a function that vanishes in the absence of spatial correlations in the system. The structure factor  $S(\mathbf{k})$  is related to the Fourier transform of  $h(\mathbf{r})$ , denoted by  $\tilde{h}(\mathbf{k})$  via the expression,

$$S(\mathbf{k}) = 1 + \rho \tilde{h}(\mathbf{k}), \quad (6)$$

which is directly measurable via the scattering intensity.<sup>72</sup>

We note that the full distribution  $P[N(R)]$  and its associated moments  $\sigma_N^2(R)$  and  $\gamma_1(R)$  are directly related to the many-body correlation functions  $g_n$  for  $n \geq 2$ , as shown in the study by Torquato, Kim, and Klatt.<sup>11</sup> For example, the number variance  $\sigma_N^2(R)$  can be computed via the two-body correlation function,

$$\sigma_N^2(R) = \langle N(R) \rangle \left[ 1 + \rho \int_{\mathbb{R}^d} h(\mathbf{r}) \alpha_2(r; R) d\mathbf{r} \right], \quad (7)$$

where  $\alpha_2(r; R)$  is the intersection volume of two spherical windows of radius  $R$ , scaled by the sphere volume  $v_1(R)$ , whose centers are separated by the distance  $r$ . Similarly,  $\gamma_1(R)$  is related to  $g_2$  and  $g_3$ , and  $\gamma_2(R)$  is related to  $g_2$ ,  $g_3$ , and  $g_4$ ; see Appendix B.

The probability distribution  $P[N(R) = m]$  can be expressed as a series expansion involving  $g_n$  for  $n = 2, 3, \dots$ <sup>73,74</sup> Truncating such series at the two- and three-body levels yields sharp bounds on  $P[N(R) = m]$ . In particular, we obtained two- and three-body bounds on the void probability  $P[N(R) = 0]$ , respectively, as follows:<sup>11</sup>

$$P[N(R) = 0] \leq [1 - \langle N(R) \rangle] \left[ 1 - \frac{\langle N(R) \rangle}{2} \right] + \frac{\sigma_N^2(R)}{2}, \quad (8)$$

$$P[N(R) = 0] \geq [1 - \langle N(R) \rangle] \left[ 1 - \frac{\langle N(R) \rangle}{2} \right] \left[ 1 - \frac{\langle N(R) \rangle}{3} \right] + \left[ 1 - \frac{\langle N(R) \rangle}{2} \right] \sigma_N^2(R) - \frac{\sigma^3(R)}{6} \gamma_1(R). \quad (9)$$

### 3. Tetrahedral order parameter $q$

We characterize the degree of tetrahedrality of each of our state points by the tetrahedral order parameter  $q$  from Errington and Debenedetti<sup>59</sup> that was motivated by Chau and Hardwick.<sup>75</sup> This quantity is defined as

$$q := 1 - \frac{3}{8} \sum_{j=1}^3 \sum_{k=j+1}^4 \left( \frac{1}{3} + \cos \psi_{jk} \right)^2, \quad (10)$$

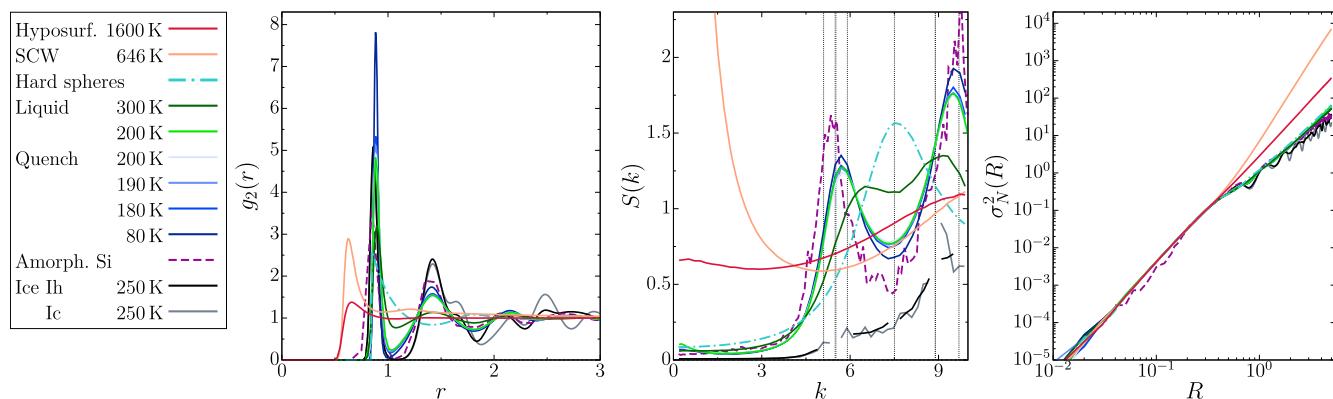
where  $\psi_{jk}$  is the angle formed by the lines joining the oxygen atom of a given molecule to those of its (nearest) neighbors  $j$  and  $k$ . Equation (10) is rescaled so that its average  $\langle q \rangle$  is 0 for an ideal gas without any orientational order and 1 for a perfect tetrahedral network.

## III. RESULTS

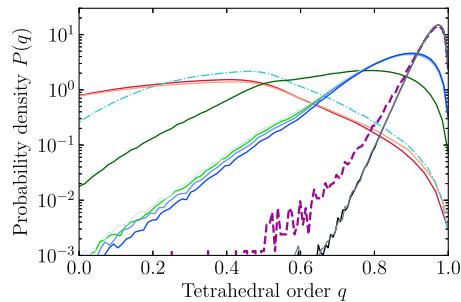
We begin with comparisons of classical two-body statistics and distributions of  $q$  for all of our states. Strong differences in  $\langle q \rangle$  lead to only a few distinguishing features in the two-body statistics. Then, we turn to the higher-order moments with clear features of tetrahedrality that point toward a non-monotonic temperature dependence of the speed of convergence to a Gaussian number distribution. This non-monotonicity is confirmed via  $l_2(R)$ . Its direct physical implications are discussed in the exemplary case of hydrophobic solubility.

Figure 2 shows the two-body statistics for all the states considered here, i.e., the pair correlation function  $g_2(r)$  in real space, the corresponding structure factor  $S(k)$  in Fourier space, and the number variance  $\sigma_N^2(R)$  that directly quantifies density fluctuations at the two-body level. The pair correlation function of equilibrium water at room temperature compared to those of the nonequilibrium quenches (down to 180 K) differ primarily in the height of the first two peaks and the depth of the first minimum (as a function of the radial distance  $r$ ), indicating the development of more structured local environments as the temperature decreases. The ice phases clearly exhibit Bragg peaks in  $S(k)$  (dotted lines), but at large scales, their number variance  $\sigma_N^2(R)$  has the same scaling qualitatively as the other non-hyperuniform phases. At the critical point, the state of water becomes anti-hyperuniform, i.e.,  $\sigma_N^2(R)$  grows faster than the volume, and  $S(k)$  diverges at the origin. In contrast, at high temperatures (above 1600 K), water becomes hypersurfacial, i.e., the surface term in the scaling of the number variance  $\sigma_N^2(R)$  vanishes,<sup>13</sup> i.e.,  $\sigma_N^2(R)$  exhibits a volume-like scaling even for local fluctuations.

Figure 3 compares the probability density  $P(q)$  of the tetrahedral order parameter  $q$  for all of our states. As expected, the ice phases and the amorphous silicon have the highest degree of tetrahedrality, followed by the quenches (the higher the final temperature, the lower the degree of tetrahedral order) and finally the equilibrium water at  $T = 300$  K. The high-temperature phases exhibit no explicit tetrahedral order, neither does the hard-sphere liquid.



**FIG. 2.** Two-body statistics for all states considered here: (a) pair correlation function  $g_2(r)$ , (b) structure factor  $S(k)$ , and (c) number variance  $\sigma_N^2(R)$ . In the legend, the states are sorted such that their mean values of tetrahedrality  $\langle q \rangle$  increase from the top to bottom. As noted in Sec. II B, all the quantities are estimated at unit number density  $\rho = 1$ .

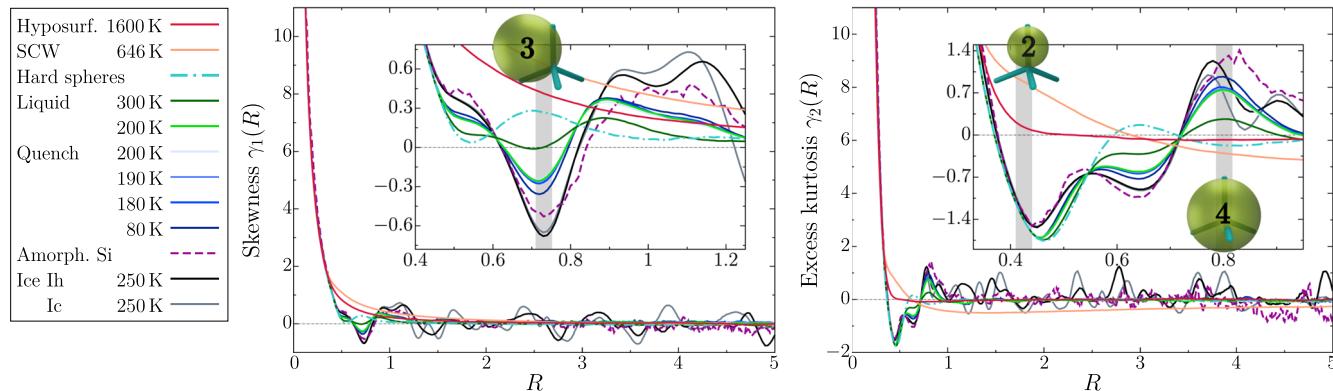


**FIG. 3.** Semi-log plot of the probability density  $P(q)$  of the tetrahedral order parameter  $q$  for all the states considered here; for the legend, see Fig. 2.

### A. Signatures of tetrahedrality in higher-order moments

We begin our analysis of the higher-order density fluctuations by computing the third- and fourth-order central moments  $\gamma_1(R)$  and  $\gamma_2(R)$  of the number distribution, which embody up to four-body correlations; see Eqs. (22) and (23) in the study by Torquato, Kim, and Klatt.<sup>11</sup> Thus, our analysis reveals non-trivial higher-order correlations in all states of water, including explicit features of tetrahedrality.

Figure 4 shows the skewness  $\gamma_1(R)$  and excess kurtosis  $\gamma_2(R)$  of the number distributions of oxygen atoms in a spherical observation window of radius  $R$ . At specific radii,  $\gamma_1(R)$  and  $\gamma_2(R)$  exhibit salient features that clearly scale with the degree of tetrahedral order for all our states.



**FIG. 4.** Skewness  $\gamma_1$  (left) and excess kurtosis  $\gamma_2$  (right) associated with window radius  $R$  are compared for all the water states considered here, amorphous silicon with a high degree of tetrahedral order, and a hard-sphere liquid with a low degree of tetrahedral order. At three specific radii (shown in the gray shades), the features in  $\gamma_1$  and  $\gamma_2$  consistently scale with the degree of tetrahedral order in these states. At these radii, the possible number of atoms per observation window increases by one for a regular tetrahedral node. These values are indicated in the attached figures of the tetrahedral nodes and spherical observation windows.

Some of these radii directly correspond to characteristic distances in a regular tetrahedral network. To explain this relation, we first have to choose an appropriate bond length of the tetrahedral network, e.g., here, we choose the distance to the first peak of the pair correlation function, as shown in Fig. 2. Then, the gray band at  $R \approx 0.73$  in the plot of  $\gamma_1(R)$  corresponds to the smallest radius for which the spherical observation window can contain three atoms. Close to this radius, we can, therefore, expect that the higher the degree of tetrahedrality, the lower the probability of finding four or five atoms within this radius and, hence the more left-tailed  $P[N(R)]$  will be, i.e., the more negative will be the value of  $\gamma_1(R = 0.73)$ . This scaling of  $\gamma_1(R = 0.73)$  with the degree of tetrahedrality is quantitatively confirmed in Fig. 5(a), which plots  $\gamma_1(R = 0.73)$  against the average tetrahedral order parameter  $\langle q \rangle$ .

Attached to the gray band at  $R \approx 0.73$  shown in the inset of Fig. 4 is a small figure that indicates the tetrahedral node and the spherical observation window, which contains up to three atoms of the network. As we have seen, this borderline case for three atoms imprints a clear feature of tetrahedrality on the skewness  $\gamma_1(R)$ , i.e., the third-order moment of the density fluctuations.

The excess kurtosis  $\gamma_2(R)$ , i.e., the fourth-order moment, exhibits a similar feature at the smallest radius for which the observation window can contain four atoms of the regular tetrahedral network. This radius is  $R \approx 0.80$ , and the scaling is confirmed in Fig. 5(b). The values for the ice phases are slightly off, which can be explained via Fig. 4. Close to  $R \approx 0.80$ ,  $\gamma_2(R)$  exhibits a local maximum as a function of  $R$  for all our states except at high temperatures. However, the peak positions of the ice phases are slightly shifted with respect to those of the disordered water phases. The latter agrees with the peak of amorphous silicon. The different peak positions for the ice phases are indicative of different distortions in the tetrahedral networks compared to a regular network (e.g., different ring statistics and consequently bending angles).

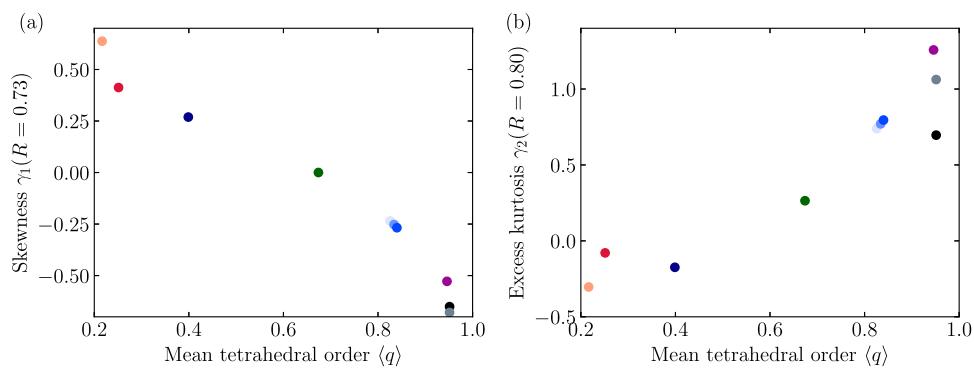
At  $R = 0.42$ , i.e., the smallest possible radius of a window with two atoms of our regular tetrahedral network, we find that  $\gamma_1(R)$  and  $\gamma_2(R)$  are still dominated by the large void probability, i.e., the probability that the window contains no atoms. Hence, features that scale with the degree of tetrahedrality only emerge at slightly larger radii of about 0.45–0.50.

Together, these features of  $\gamma_1(R)$  and  $\gamma_2(R)$  offer a refined and statistically robust characterization of the tetrahedral order in disordered phases of water both in and out of equilibrium—with immediate physical implications. In fact, the additional structural information reveals a behavior in the higher-order moments that is non-monotonic with respect to temperature.

In the search for a physical explanation of this non-monotonic behavior, we compare  $\gamma_1(R)$  and  $\gamma_2(R)$  of our water states to those of a simple liquid without tetrahedral order, namely, the hard-sphere liquid. We first focus on  $R \approx 0.73$ , where  $\gamma_1(R)$  exhibits the feature of tetrahedrality discussed above, i.e., the local minimum as a function of  $R$  that becomes more negative for more tetrahedrally ordered water phases. In contrast,  $\gamma_1(R)$  of the hard-sphere liquid has a local maximum close to this radius. Its functional value, however, is still smaller than those of the high-temperature water states that are dominated by thermal fluctuations.

Hence, we observe that the stronger the thermal fluctuations are, the larger is  $\gamma_1(R = 0.73)$ , but the stronger the tetrahedral order is, the smaller is  $\gamma_1(R = 0.73)$ . At room temperature, these two influences approximately cancel out so that  $\gamma_1(R = 0.73) \approx 0$ . Similar effects can be observed for larger radii and for  $\gamma_2(R)$ . Overall,  $\gamma_1(R)$  and  $\gamma_2(R)$  appear to converge to zero more quickly for water at room temperature than for our other state points.

Importantly, these temperature non-monotonicities that we have found in the higher-order moments complement the known non-monotonic temperature dependencies of thermodynamic response functions (isothermal compressibility and heat capacity), which anomalously increase upon cooling. Interestingly, these anomalous temperature dependencies of the thermodynamic response functions are not strongly reflected in the corresponding pair correlation functions, which show only a mild increase in the strength of correlations as temperature decreases (see Fig. 2). By contrast, our higher-order moments reveal distinct structural signatures, such as the pronounced features in  $\gamma_1(R)$  and  $\gamma_2(R)$  at specific radii. The functional values at these radii change with temperature, and their absolute values are minimal close to room temperature, hence mirroring the non-monotonic trends in the response functions as functions of temperature.



**FIG. 5.** For all our states, the mean tetrahedral order  $\langle q \rangle$  is plotted against the (a) skewness and (b) excess kurtosis at the characteristic radii. The states are represented by the same color code as that shown in Fig. 4.

In Sec. III B, we use the Gaussian distance metric to quantify how quickly the higher-order moments tend toward zero. Thus, we confirm that the speed of convergence has a non-monotonic temperature dependence (as surmised above), and specifically, the convergence is faster at room temperature than at higher or lower temperatures. Then, we will discuss the chemical–physical implications for hydrophobic solubility and beyond.

## B. Effect of tetrahedrality on the Gaussian distance metric

The combined effect of all higher-order moments on the number distribution can be best quantified by the Gaussian distance metric  $l_2(R)$ . Figure 6 compares our results for all the states. We observe that  $l_2(R)$  decays at large radii for all disordered states of water (i.e., for all states except for the two ice phases). This observation suggests that a central limit theorem (CLT) holds for large  $R$  (and thus correspondingly large  $N$ ) with respect to the chosen metric.

Since any CLT depends on its metric, our proposed CLT still agrees with the heavy left tails reported for very large radii.<sup>42</sup> There, indirect umbrella sampling<sup>37,41</sup> revealed heavy left tails of  $P[N(R)]$ , which can be related to effects at liquid–vapor interfaces.<sup>42</sup> However, since the onset of the tails appears to converge to zero in the thermodynamic limit, they are consistent with a CLT in an  $l_2$  metric. Note that these heavy left tails in water are distinct from the more commonly observed right tails for random point processes as reported for a broad spectrum of models in the study by Torquato, Kim, and Klatt.<sup>11</sup>

The CLT and its speed of convergence for a certain range of radii indicates how well  $P[N(R)]$  can be approximated by a

Gaussian distribution and hence by two-body statistics. As an interesting side remark, the renormalized Gaussian approximation of Ashbaugh, Vats, and Garde<sup>40</sup> happens to coincide with the reference distribution of the Gaussian distance metric  $l_2(R)$ .<sup>11</sup>

For one relatively large radius and one relatively small radius, we also show the corresponding number distributions in Fig. 6. Already at the second peak of  $l_2(R)$ , the Gaussian distance metric is smaller for water at room temperature than under non-ambient conditions, whether it is warmer or colder. This non-monotonic temperature dependence essentially continues for larger radii and indeed confirms within our accuracy that among all states considered, the number distribution converges fastest to the normal distribution for equilibrium water at room temperature.

At high temperatures, thermal fluctuations enhance density fluctuations, which leads to positively skewed number distributions at a given length scale  $R$ , while at low temperatures, the emergence of tetrahedral order increases the correlation length and leads to negatively skewed number distributions at the same length scale  $R$ . Under ambient conditions, those two trends balance out, which results in a fast convergence of the number distribution to a Gaussian distribution.

As noted above, this balance between thermal fluctuations and tetrahedral order is similar to the surmised reason for the anomalous minimum in the isothermal compressibility upon cooling. In our case, however, the non-monotonic temperature dependence already appears at a local scale, applies to non-equilibrium phases, and it explicitly affects the higher-order moments with immediate physical implications, e.g., on the hydrophobic solubility, as discussed in Sec. III C. We will broaden the scope of the discussion in Sec. IV.

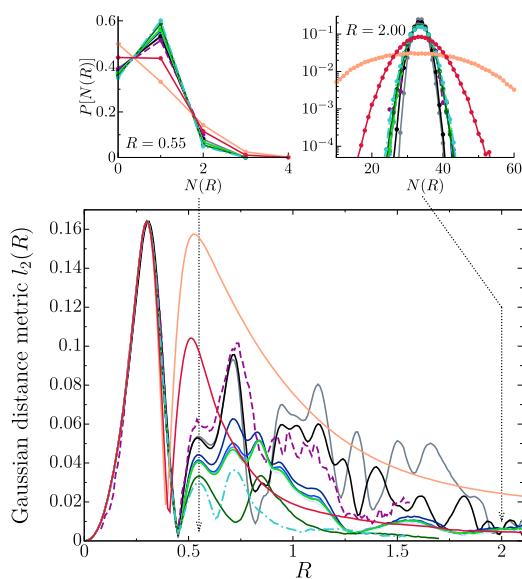
## C. Implications on hydrophobic solubility

A prominent example of the chemical–physical importance of density fluctuations in water is the solubility of hydrophobic solutes in water. Specifically, the excess chemical potential  $\mu^{\text{ex}}$  of hydration for a hard-particle solute of radius  $R$  can be expressed by the probability of the random formation of a spherical cavity in water, i.e., the void probability  $P[N(R) = 0]$ ,<sup>76,77</sup>

$$\mu^{\text{ex}} = -k_B T \ln P[N(R) = 0]. \quad (11)$$

The estimation of  $P[N(R) = 0]$  for liquid water has been a particularly productive field of research starting with a seminal paper by Hummer *et al.*<sup>23</sup> and subsequent studies over the past two decades.<sup>34–42</sup>

Hummer *et al.*<sup>23</sup> estimated  $P[N(R) = 0]$  via an information-theoretic approximation, i.e., the probability distribution  $\hat{P}[N(R)]$  that maximizes the cross entropy (relative to a flat prior distribution) given the first  $M$  moments of  $P[N(R)]$ . Using the first two moments of  $P[N(R)]$ , Hummer *et al.* obtained (at ambient temperatures) a surprisingly good approximation of the void probability for small to intermediate radii, i.e., up to about 10 Å.<sup>42</sup> Counterintuitively, the accuracy of the information-theoretic approximation was found to deteriorate if higher-order moments were taken into account, i.e.,  $2 < M < 7$ , as shown in the study by Hummer *et al.*<sup>58</sup>



**FIG. 6.** Gaussian distance metric  $l_2(R)$  as a function of window radius  $R$ ; for the legend, see Fig. 4. The insets show the number distributions at two different radii, indicated by the dotted arrows.

This surprising accuracy at near-ambient conditions of the information-theoretic approximation based on only the first two moments evokes at least three questions:

- Why does the information-theoretic approximation, counter-intuitively, worsen when higher-order moments are included?
- Does the information-theoretic approximation apply to water in out-of-equilibrium or metastable states or solid phases?
- Should the tetrahedrality of water impact the accuracy of the prediction of the solubility?

To address these three questions, we compute (from our simulations) the void probability  $P[N(R) = 0]$ , also called cavity formation probability. Then, we determine the information-theoretic approximation with varying numbers of constraints  $M$  (always employing the flat prior probability as in Hummer *et al.*<sup>23</sup>). Finally, Fig. 7 compares our simulations results for  $P[N(R) = 0]$  (dots) to the information-theoretic approximation (colored lines).

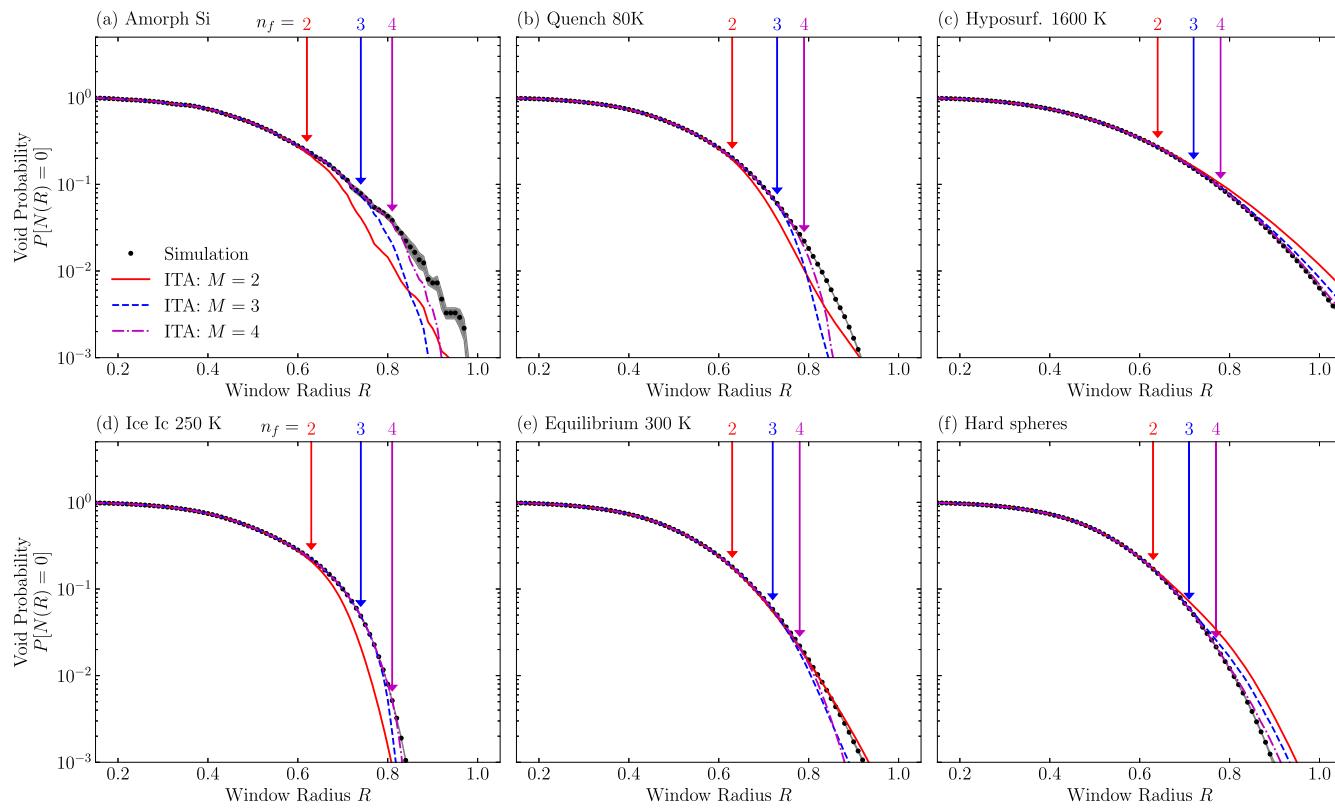
While we confirm for water at ambient temperatures, the superiority of the information-theoretic approximation with  $M = 2$  over those with higher-order moments, we find that it is an exception. For all other states, the information-theoretic approximation is accurate only for a range of radii where the number of constrained moments essentially fix the entire probability distribution; the higher the

number of moments, the larger this range of radii, which can be quantified by the following rule of thumb.

The information-theoretic approximation with  $M$  constraints is accurate if there are, at most,  $M$  non-negligible values of  $P[N(R)]$ . In that case, the  $M$  moments virtually specify all probabilities. More specifically, we define the effective support size  $n_f(R)$  as the minimal number  $n$  for which  $\sum_{m=n+1}^{\infty} P[N(R) = m] < P[N(R) = 0]$ . Then, the information-theoretic approximation with  $M$  constraints is accurate for a range of radii where  $n_f(R) \leq M$ , as shown in Fig. 7.

For the same range of radii, we also obtain accurate predictions via the rigorous bounds from Torquato, Kim, and Klatt.<sup>11</sup> The upper and lower bounds virtually coincide within this regime, which may facilitate an analytic approach to hydrophobic solubility for small radii. For larger radii, the information-theoretic approximation prediction is more precise even though it is not perfect.

At these radii, slightly larger than the constrained range, the deviations of the information-theoretic approximation from our simulation data display a consistent pattern. The information-theoretic approximation underestimates  $P[N(R) = 0]$  when the skewness is negative, which corresponds to a tetrahedral order higher than that for water at ambient temperatures, and the



**FIG. 7.** Comparison of the void probability  $P[N(R) = 0]$  from simulations and the information-theoretic approximation—varying the highest order of constrained moments  $M = 2, 3, 4$ . The vertical arrows indicate the upper bounds on the effective support sizes at three different values ( $n_f = 2, 3, 4$ ); they coincide with the radii at which the predictions begin to deviate from the simulations.

information-theoretic approximation overestimates  $P[N(R) = 0]$  when the skewness is positive, which corresponds to a low tetrahedral order. For water at room temperature, the skewness and excess kurtosis almost vanish, as discussed above, resulting in accurate Gaussian predictions of the information-theoretic approximation with  $M = 2$ . Hence, the surprisingly good performance of the information-theoretic approximation under ambient conditions appears to be again related, via  $\gamma_1(R)$ , to a balance of thermal fluctuations and tetrahedral order.

In answer to the questions above, (ii) the information-theoretic approximation is a valuable approximation for all our states, but a similar accuracy as for water at room temperature is obtained only for small radii, where the first  $M$  moments fix the distribution. (iii) Both tetrahedrality and thermal fluctuations impact the accuracy of the information-theoretic approximation but with opposing effects, i.e., underestimation or overestimation of  $P[N(R) = 0]$ , respectively. (i) At room temperature, these effects roughly cancel out. For all our other states, higher-order moments improve the information-theoretic approximation; more precisely, the range of radii for which we obtain accurate predictions increases. For example, if we constrain four instead of two moments, the range of radii with accurate predictions increases from about 0.6 to 0.8.

#### IV. DISCUSSION

We have studied here the link between physical properties and density fluctuations in water by going beyond the two-body level to higher-order moments and by comparing a great variety of states of water across a broad range of temperatures 80–1600 K. These states include ice phases, equilibrium liquid water, supercritical water, and disordered nonequilibrium quenches, and we compared them to two further reference systems: equilibrium hard spheres, representing a simple liquid, and a continuous random network, representing amorphous silicon. We analyzed all of our samples with a recently developed, advanced platform for local density fluctuations.<sup>11</sup> This approach includes robust estimates of higher-order moments (which enables us to capture crucial information about  $n$ -body correlations) and a Gaussian distance metric.

Our analysis reveals how water's tetrahedral order affects not only the number variance but also the higher-order moments of local density fluctuations since tetrahedrality is a many-body property. Specifically, we observe that the third- and fourth-order central moments,  $\gamma_1(R)$  and  $\gamma_2(R)$ , scale with the mean tetrahedral order parameter  $\langle q \rangle$  at two characteristic length scales  $R = 0.73$  and  $0.80$ , respectively. This scaling clearly indicates that the skewness and excess kurtosis entail signatures of tetrahedrality on the higher-order correlations for all our phases. Moreover, the corresponding radii can be directly related to characteristic distances in the tetrahedral networks.

The Gaussian distance metric  $l_2(R)$  as a function of the radius  $R$  has a local minimum at  $R \approx 0.72$  for liquid water at  $T = 300$  K but a local maximum for the quenches and the supercooled liquid at  $T = 200$  K. Although the pair correlation functions appear to be similar for all our liquid-like states of water at a wide range of conditions, the higher-order moments reveal distinct structural differences, e.g., related to the degree of tetrahedrality or to how

close the density fluctuations are to Gaussian. While such structural differences can easily be missed by two-body characteristics, they are clearly captured by the higher-order moments. Hence, our results further motivate the need for experimental methods to ascertain three- and higher-body correlations in water systems, e.g., via isothermal pressure derivatives of the structure factor or a spherical harmonic analysis.<sup>60–62</sup>

A key insight into our higher-order moment analysis is the newly found temperature non-monotonicity in the Gaussian distance metric  $l_2(R)$ . First of all, the convergence or non-convergence to a Gaussian distribution distinguishes the disordered from the ordered phases. Specifically, in contrast to the two crystalline ice phases, the distributions of all our disordered states become close to a Gaussian for large radii as measured by  $l_2(R)$ . For these liquid-like states, we observe that under ambient conditions, tetrahedral order and thermal fluctuations balance out and hence  $l_2(R)$  converges to zero most rapidly. The convergence slows down at higher and lower temperatures; in the former case, because of larger thermal fluctuations and in the latter case, because of higher tetrahedral order.

This non-monotonicity distinguishes itself from the well-known anomaly in the isothermal compressibility in at least two aspects. Even though the isothermal compressibility can be related to density fluctuations, more precisely, the asymptotic number variance, it only holds in the limit of infinite radii and only pertains to the second moment. In our case, we find that tetrahedrality also induces non-monotonic temperature dependencies locally for the higher-order moments.

Another intriguing aspect of water's anomalies is the nested structural, dynamic, and thermodynamic anomalies noted by Errington and Debenedetti.<sup>59</sup> Specifically, they observe a region of “structural” anomalies, where water's translational and orientational order decrease upon compression; a region of “dynamic” anomalies, where water's diffusion coefficient increases upon compression; and a region of “thermodynamic” anomalies, where density decreases upon cooling at constant pressure. The set of temperatures and densities that define the region of structural anomalies completely contains the region of dynamic anomalies, which, in turn, contains the region of thermodynamic anomalies. Given that our present approach provides a platform to characterize higher-order structural information about water at all length scales, in future work, it would be interesting to study how the skewness, kurtosis, and Gaussian distance metric might behave at state points in the vicinity of water's nested anomalies. Such an effort may reveal further structural links between water's anomalous local, mesoscale, and macroscale phenomena.<sup>59</sup>

A consequence of this non-monotonic behavior is that higher-order moments are no longer negligible relative to the first and second moments once we consider water states away from ambient conditions. Hence, second-order approximations that assume Gaussian density fluctuations and that work well close to room temperature will become less accurate at both high and low temperatures. We demonstrate this effect for the information-theoretic approximation of hydrophobic solubility.<sup>23</sup> With the prominent exception of water under ambient conditions, we find that the information-theoretic approximation is accurate only in a range where the constrained moments essentially determine the entire

probability distribution. Therefore, the higher-order moments generally improve the approximation for the larger solute radii as one departs from ambient conditions.

Moreover, for radii beyond the range where the first two moments provide accurate predictions, the sign of  $\gamma_1(R)$  indicates whether the information-theoretic approximation over- or underestimates the hydrophobic solubility for our states. The sign of  $\gamma_1(R)$ , in turn, is linked to the question whether tetrahedrality or thermal fluctuations have a greater influence on the local degree of order and disorder. Thus, our higher-order moment analysis not only improved the predictions quantitatively and provided estimates for the range where the approximation is accurate and which deviations are to be expected but also the higher-order moments provided us with physical insights into what determines the local structural features of a wide variety of water phases.

One of the most immediate applications of hydrophobic solubility is in biology, which of course is typically restricted to near-ambient conditions. Thus, the utility of the information-theoretic approximation even with  $M = 2$  is readily apparent. However, there are many technologically and scientifically impactful cases where predicting water's interactions with hydrophobic solutes may also be important under conditions far from ambient, such as for understanding the impact of hydrophobic aerosol particles in the atmosphere<sup>51,52</sup> (i.e., low temperature), or for understanding life in exotic environments such as on astronomical bodies, e.g., comets, planets<sup>53</sup> (i.e., low-temperature, low-pressure), or near hydrothermal vents<sup>54</sup> (i.e., high-temperature, high-pressure). In such cases, considering higher-order moments of  $P[N(R)]$  may prove useful.

Finally, we point out that the higher-order moments generically provide a robust characterization of  $n$ -body correlations in water's tetrahedral network. Such additional structural information can, for example, enhance or complement recent unsupervised machine learning approaches to classify high- and low-density structures in liquid water.<sup>78</sup> One could also envision combining these structural characterization techniques with recent advances in machine-learned interaction potentials<sup>79–81</sup> to explore structure across length scales as derived from *ab initio* models. More broadly, our findings could also generalize to other highly polar (e.g., ammonia and HF) and/or network-forming liquids (e.g., Si and Ge).<sup>82</sup>

## ACKNOWLEDGMENTS

We thank G. T. Barkema and N. Mousseau for providing their samples of CRNs. This work was supported in part by the National Science Foundation Grant No. CBET-2133179 and the U.S. Army Research Office under Cooperative Agreement No. W911NF-22-2-0103. M.A.K. acknowledges funding and support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through SPP 2265 under Grant Nos. KL 3391/2-2, WI 5527/1-1, and LO 418/25-1, as well as by the Initiative and Networking Fund of the Helmholtz Association through the Project "DataMat." This work used computational resources managed by Princeton Research Computing, a consortium of groups including the Princeton Institute for Computational Science and Engineering (PICSciE) and the Office of Information Technology's High Performance Computing Center and Visualization Laboratory at Princeton University.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

## Author Contributions

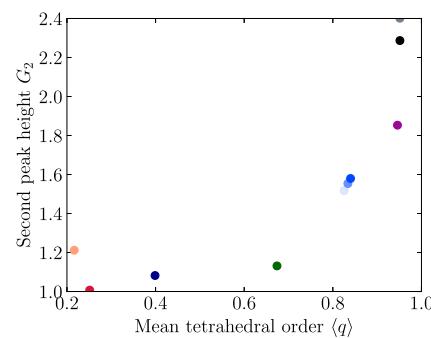
**Michael A. Klatt:** Data curation (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Jaeuk Kim:** Data curation (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Thomas E. Gartner III:** Data curation (equal); Investigation (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Salvatore Torquato:** Conceptualization (equal); Investigation (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## APPENDIX A: CORRELATION OF THE SECOND PEAK OF $g_2(r)$ WITH THE TETRAHEDRAL ORDER

The value of the two-body correlation function  $g_2(r)$  at the second peak, which is denoted by  $G_2$  here, has been employed by Sellberg *et al.*<sup>48</sup> to quantify the degree of tetrahedral order of liquid water. For purposes of comparison, we apply this parameter to all states considered here. Figure 8 clearly shows a positive correlation between  $G_2$  and the mean tetrahedral order ( $q$ ) for all the states of water considered in this paper (and similar to the supplementary Fig. S21 from Sellberg *et al.*<sup>48</sup> in a different representation for temperatures 200–340 K).



**FIG. 8.** The two-body measure  $G_2$ , which is the height of the second peak of  $g_2(r)$ , is plotted as a function of the mean tetrahedral order ( $q$ ) for all the states considered here. These states are represented by the same color code as shown in Fig. 2.

APPENDIX B: FORMULAS FOR THE SKEWNESS  
AND EXCESS KURTOSIS IN TERMS OF *N*-BODY  
CORRELATION FUNCTIONS

Here, we represent the expressions for the skewness  $\gamma_1(R)$  and excess kurtosis  $\gamma_2(R)$  associated with the spherical observation window of radius  $R$  in terms of the *n*-body correlation functions  $g_n(\mathbf{r}^n)$ ; see the study by Torquato, Kim, and Klatt<sup>11</sup> for derivations. The skewness, defined in Eq. (2), can be written as

$$\begin{aligned} \gamma_1(R) = & [\sigma_N^2(R)]^{-3/2} \left\{ \rho v_1(R) + 3\rho^2 \int_{\mathbb{R}^d} h(\mathbf{r}) v_2^{\text{int}}(r; R) d\mathbf{r} \right. \\ & + \rho^3 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [g_3(\mathbf{r}^3) - 3g_2(\mathbf{r}_{12}) + 2] \\ & \times v_3^{\text{int}}(\mathbf{r}^3; R) d\mathbf{r}_2 d\mathbf{r}_3 \left. \right\}, \end{aligned} \quad (\text{B1})$$

where  $v_n^{\text{int}}(\mathbf{r}^n; R)$  is the intersection volume of *n* spheres of radius  $R$  centered at positions  $\mathbf{r}^n = \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$ . We see that  $\gamma_1(R)$  encodes up to three-body information. The excess kurtosis, defined in Eq. (3), can be written as

$$\begin{aligned} \gamma_2(R) = & [\sigma_N^2(R)]^{-2} \left\{ \rho v_1(R) + 7\rho^2 \int_{\mathbb{R}^d} h(\mathbf{r}) v_2^{\text{int}}(r; R) d\mathbf{r} \right. \\ & + 6\rho^3 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [g_3(\mathbf{r}^3) - 3g_2(\mathbf{r}_{12}) + 2] \\ & \times v_3^{\text{int}}(\mathbf{r}^3; R) d\mathbf{r}_2 d\mathbf{r}_3 \\ & + \rho^4 \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} [g_4(\mathbf{r}^4) - 4g_3(\mathbf{r}^3) + 12g_2(\mathbf{r}_{12}) - 6] \\ & \times v_4^{\text{int}}(\mathbf{r}^4; R) d\mathbf{r}_2 d\mathbf{r}_3 d\mathbf{r}_4 \\ & \left. - 3 \left[ \rho^2 \int_{\mathbb{R}^d} g_2(\mathbf{r}) v_2^{\text{int}}(r; R) d\mathbf{r} \right]^2 \right\}. \end{aligned} \quad (\text{B2})$$

We see that  $\gamma_2(R)$  encodes up to four-body information.

## REFERENCES

- 1 P. Schofield, *Proc. Phys. Soc.* **88**, 149 (1966).
- 2 F. Carmona and P. Delhaes, *J. Appl. Phys.* **49**, 618 (1978).
- 3 J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids* (Academic Press, New York, 1986).
- 4 K. Jørgensen, J. H. Ipsen, O. G. Mouritsen, D. Bennett, and M. J. Zuckermann, *Biochim. Biophys. Acta, Biomembr.* **1067**, 241 (1991).
- 5 P. J. E. Peebles, *Principles of Physical Cosmology* (Princeton University Press, Princeton, 1993).
- 6 S. Torquato, *Int. J. Solids Struct.* **37**, 411 (2000).
- 7 A. C. Lavery, R. W. Schmitt, and T. K. Stanton, *J. Acoust. Soc. Am.* **114**, 2685 (2003).
- 8 S. Torquato, A. Scardicchio, and C. E. Zachary, *J. Stat. Mech.: Theory Exp.* **2008**, P11019.
- 9 M. A. Klatt, G. Last, and N. Henze, "A genuine test for hyperuniformity," *arXiv:2210.12790* [cond-mat, stat] (2022).
- 10 S. Torquato and F. H. Stillinger, *Phys. Rev. E* **68**, 041113 (2003).
- 11 S. Torquato, J. Kim, and M. A. Klatt, *Phys. Rev. X* **11**, 021028 (2021).
- 12 Y. Zheng, A. D. S. Parmar, and M. Pica Ciamarra, *Phys. Rev. Lett.* **126**, 118003 (2021).
- 13 S. Torquato, *Phys. Rep.* **745**, 1 (2018).
- 14 S. Torquato and J. Kim, *Phys. Rev. X* **11**, 021002 (2021).
- 15 K. Vynck, R. Pierrat, R. Carminati, L. S. Froufe-Pérez, F. Scheffold, R. Sapienza, S. Vignolini, and J. J. Sáenz, *Rev. Mod. Phys.* **95**, 045003 (2023).
- 16 Y. Xu *et al.*, *Phys. Rev. E* **96**, 043301 (2017).
- 17 S. Torquato, *Adv. Water Resour.* **140**, 103565 (2020).
- 18 M. A. Klatt, R. M. Ziff, and S. Torquato, *Phys. Rev. E* **104**, 014127 (2021).
- 19 M. A. Klatt, P. J. Steinhardt, and S. Torquato, *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2213633119 (2022).
- 20 Y. Jiao, T. Lau, H. Hatzikirou, M. Meyer-Hermann, J. C. Corbo, and S. Torquato, *Phys. Rev. E* **89**, 022721 (2014).
- 21 P. A. Monderkamp, F. J. Schwarzlindahl, M. A. Klatt, and H. Löwen, *Mach. Learn.: Sci. Technol.* **3**, 045024 (2022).
- 22 D. Chandler, *Phys. Rev. E* **48**, 2898 (1993).
- 23 G. Hummer, S. Garde, A. E. García, A. Pohorille, and L. R. Pratt, "An information theory model of hydrophobic interactions," *Proc. Natl. Acad. Sci.* **93**(17), 8951–8955 (1996).
- 24 S. Garde, G. Hummer, A. E. García, M. E. Paulaitis, and L. R. Pratt, *Phys. Rev. Lett.* **77**, 4966 (1996).
- 25 K. Lum, D. Chandler, and J. D. Weeks, *J. Phys. Chem. B* **103**, 4570 (1999).
- 26 L. R. Pratt, *Annu. Rev. Phys. Chem.* **53**, 409 (2002).
- 27 P. R. t. Wolde, *J. Phys.: Condens. Matter* **14**, 9445 (2002).
- 28 D. Chandler, *Nature* **437**, 640 (2005).
- 29 L. R. Pratt and D. Chandler, *J. Chem. Phys.* **67**, 3683 (1977).
- 30 D. N. LeBard and D. V. Matyushov, *Phys. Rev. E* **78**, 061901 (2008).
- 31 J. Mittal and G. Hummer, *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20130 (2008).
- 32 J. Russo, K. Akahane, and H. Tanaka, *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3333 (2018).
- 33 P. G. Debenedetti, *J. Phys.: Condens. Matter* **15**, R1669 (2003).
- 34 R. Godawat, S. N. Jamadagni, and S. Garde, *Proc. Natl. Acad. Sci. U. S. A.* **106**, 15119 (2009).
- 35 A. J. Patel, P. Varilly, and D. Chandler, *J. Phys. Chem. B* **114**, 1632 (2010).
- 36 A. J. Patel, P. Varilly, S. N. Jamadagni, H. Acharya, S. Garde, and D. Chandler, *Proc. Natl. Acad. Sci. U. S. A.* **108**, 17678 (2011).
- 37 A. J. Patel, P. Varilly, D. Chandler, and S. Garde, *J. Stat. Phys.* **145**, 265 (2011).
- 38 R. S. Singh, J. W. Biddle, P. G. Debenedetti, and M. A. Anisimov, *J. Chem. Phys.* **144**, 144504 (2016).
- 39 Z. Jiang, R. C. Remsing, N. B. Rego, and A. J. Patel, *J. Phys. Chem. B* **123**, 1650 (2019).
- 40 H. S. Ashbaugh, M. Vats, and S. Garde, *J. Phys. Chem. B* **125**, 8152 (2021).
- 41 I. Sinha, S. M. Cramer, H. S. Ashbaugh, and S. Garde, *J. Phys. Chem. B* **126**, 7604 (2022).
- 42 N. B. Rego and A. J. Patel, *Annu. Rev. Condens. Matter Phys.* **13**, 303 (2022).
- 43 F. Martelli, S. Torquato, N. Giovambattista, and R. Car, *Phys. Rev. Lett.* **119**, 136002 (2017).
- 44 F. Martelli, N. Giovambattista, S. Torquato, and R. Car, *Phys. Rev. Mater.* **2**, 075601 (2018).
- 45 T. E. Gartner, S. Torquato, R. Car, and P. G. Debenedetti, *Nat. Commun.* **12**, 3398 (2021).
- 46 M. Formanek, S. Torquato, R. Car, and F. Martelli, *J. Phys. Chem. B* **127**, 3946 (2023).
- 47 A. Nilsson and L. G. M. Pettersson, *Nat. Commun.* **6**, 8998 (2015).
- 48 J. A. Sellberg, C. Huang, T. A. McQueen, N. D. Loh, H. Laksmono, D. Schlesinger, R. G. Sierra, D. Nordlund, C. Y. Hampton, D. Starodub, D. P. DePonte, M. Beye, C. Chen, A. V. Martin, A. Barty, K. T. Wikfeldt, T. M. Weiss, C. Caronna, J. Feldkamp, L. B. Skinner, M. M. Seibert, M. Messerschmidt, G. J. Williams, S. Boutet, L. G. M. Pettersson, M. J. Bogan, and A. Nilsson, *Nature* **510**, 381 (2014).
- 49 J. L. F. Abascal and C. Vega, *J. Chem. Phys.* **123**, 234505 (2005).
- 50 L. Weng, S. L. Stott, and M. Toner, *Annu. Rev. Biomed. Eng.* **21**, 1 (2019).
- 51 A. Tabazadeh, Y. S. Djikaev, and H. Reiss, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 15873 (2002).
- 52 C. A. Yang, M. Diao, A. Gettelman, K. Zhang, J. Sun, G. McFarquhar, and W. Wu, *J. Geophys. Res.: Atmos.* **126**, e2021JD036045, <https://doi.org/10.1029/2021jd036045> (2021).
- 53 M. J. Mottl, B. T. Glazer, R. I. Kaiser, and K. J. Meech, *Geochemistry* **67**, 253 (2007).

<sup>54</sup>J. M. McDermott, J. S. Seewald, C. R. German, and S. P. Sylva, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7668 (2015).

<sup>55</sup>J.-P. Hansen and I. R. McDonald, *Theory of Simple Liquids: With Applications to Soft Matter*, 4th ed. (Academic Press, Amsterdam, 2013).

<sup>56</sup>S. Torquato, in *Random Heterogeneous Materials*, 2nd ed., *Interdisciplinary Applied Mathematics Vol. 16*, edited by S. S. Antman, L. Sirovich, J. E. Marsden, and S. Wiggins (Springer, New York, 2002).

<sup>57</sup>G. T. Barkema and N. Mousseau, *Phys. Rev. B* **62**, 4985 (2000).

<sup>58</sup>G. Hummer, S. Garde, A. E. García, M. E. Paulaitis, and L. R. Pratt, *J. Phys. Chem. B* **102**, 10469 (1998).

<sup>59</sup>J. R. Errington and P. G. Debenedetti, *Nature* **409**, 318 (2001).

<sup>60</sup>A. K. Soper, *Nucl. Instrum. Methods Phys. Res., Sect. A* **354**, 87 (1995).

<sup>61</sup>D. Dhabal, M. Singh, K. T. Wikfeldt, and C. Chakravarty, *J. Chem. Phys.* **141**, 174504 (2014).

<sup>62</sup>D. Dhabal, K. T. Wikfeldt, L. B. Skinner, C. Chakravarty, and H. K. Kashyap, *Phys. Chem. Chem. Phys.* **19**, 3265 (2017).

<sup>63</sup>C. Vega and J. L. F. Abascal, *Phys. Chem. Chem. Phys.* **13**, 19663 (2011).

<sup>64</sup>D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *J. Comput. Chem.* **26**, 1701 (2005).

<sup>65</sup>C. Vega, J. L. F. Abascal, and I. Nezbeda, *J. Chem. Phys.* **125**, 034503 (2006).

<sup>66</sup>M. Matsumoto, T. Yagasaki, and H. Tanaka, *J. Comput. Chem.* **39**, 61 (2018).

<sup>67</sup>A. Zaragoza, M. M. Conde, J. R. Espinosa, C. Valeriani, C. Vega, and E. Sanz, *J. Chem. Phys.* **143**, 134504 (2015).

<sup>68</sup>S. Torquato, *Random Heterogeneous Materials: Microstructure and Macroscopic Properties* (Springer-Verlag, New York, 2002).

<sup>69</sup>G. Jeanmairet, M. Levesque, and D. Borgis, *J. Chem. Phys.* **139**, 154101 (2013).

<sup>70</sup>J. Jin, K. S. Schweizer, and G. A. Voth, *J. Chem. Phys.* **158**, 034104 (2023).

<sup>71</sup>S. N. Chiu, D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and its Applications*, 3rd ed. (Wiley, Chichester, 2013).

<sup>72</sup>C. Kittel, *Introduction to Solid State Physics*, 8th ed. (Wiley, New York, 2005).

<sup>73</sup>D. J. Vezzetti, *J. Math. Phys.* **16**, 31 (1975).

<sup>74</sup>R. M. Ziff, *J. Math. Phys.* **18**, 1825 (1977).

<sup>75</sup>P.-L. Chau and A. J. Hardwick, *Mol. Phys.* **93**, 511 (1998).

<sup>76</sup>B. Widom, *J. Chem. Phys.* **39**, 2808 (1963).

<sup>77</sup>T. L. Beck, M. E. Paulaitis, and L. R. Pratt, *The Potential Distribution Theorem and Models of Molecular Solutions*, 1st ed. (Cambridge University Press, 2006).

<sup>78</sup>E. D. Donkor, A. Offei-Danso, A. Rodriguez, F. Sciortino, and A. Hassanal, “Beyond local structures in critical supercooled water through unsupervised learning,” *J. Phys. Chem. Lett.* **15**, 3996 (2024); arXiv:2401.16245 [cond-mat, physics:physics].

<sup>79</sup>J. Behler, *J. Chem. Phys.* **145**, 170901 (2016).

<sup>80</sup>P. Friederich, F. Häse, J. Proppe, and A. Aspuru-Guzik, *Nat. Mater.* **20**, 750 (2021).

<sup>81</sup>T. Wen, L. Zhang, H. Wang, W. E, and D. J. Srolovitz, *Mater. Futures* **1**, 022601 (2022).

<sup>82</sup>W. Hujo, B. Shadrack Jubes, V. K. Rana, C. Chakravarty, and V. Molinero, *J. Stat. Phys.* **145**, 293 (2011).