

**Disentangling structural information from core-level excitation spectra**Johannes Niskanen,<sup>1,\*</sup> Christoph J. Sahle,<sup>2</sup> Keith Gilmore,<sup>2</sup> Frank Uhlig,<sup>3</sup> Jens Smiatek,<sup>3</sup> and Alexander Föhlisch<sup>1,4</sup><sup>1</sup>*Helmholtz Zentrum Berlin für Materialien und Energie, Institute For Methods And Instrumentation For Synchrotron Radiation Research, Albert-Einstein-Strasse 15, D-12489 Berlin, Germany*<sup>2</sup>*ESRF-The European Synchrotron, 71 Avenue des Martyrs, F-38000 Grenoble, France*<sup>3</sup>*Universität Stuttgart, Institute for Computational Physics, Allmandring 3 D-70569 Stuttgart, Germany*<sup>4</sup>*Universität Potsdam, Institut für Physik und Astronomie, Karl-Liebknecht-Strasse 24/25 D-14476 Potsdam-Golm, Germany*

(Received 16 February 2017; revised manuscript received 7 July 2017; published 31 July 2017)

Core-level spectra of liquids can be difficult to interpret due to the presence of a range of local environments. We present computational methods for investigating core-level spectra based on the idea that both local structural parameters and the x-ray spectra behave as functions of the local atomic configuration around the absorbing site. We identify correlations between structural parameters and spectral intensities in defined regions of interest, using the oxygen K-edge excitation spectrum of liquid water as a test case. Our results show that this kind of analysis can find the main structure–spectral relationships of ice, liquid water, and supercritical water.

DOI: [10.1103/PhysRevE.96.013319](https://doi.org/10.1103/PhysRevE.96.013319)**I. INTRODUCTION**

At given thermodynamic conditions, atomistic systems occupy points  $\mathbf{R}$  in the configurational space with probabilistic weights  $\rho(\mathbf{R})$  that depend on the statistical ensemble [1]. Determining these probabilistic weights as a function of the thermodynamic conditions is of considerable interest. Core-level spectroscopy is sensitive to local atomic and electronic structure, in part, because the transition matrix elements depend on the spatial overlap of the strongly localized core level with the valence levels. Because distant atoms typically have only a weak effect on core spectra [2], the most significant effects on spectral shape originate from the closest neighbors surrounding the excited site. These nearest neighbors, together with the excited atom, define a local configuration. One expects to find strong correlations between this local configuration, classified in terms of selected structural parameters, and the excitation spectrum. Developing methodology to identify these correlations is the subject of this work.

Electronic transitions are not always as easy to interpret as to measure. A typical analysis technique of experimental x-ray spectra involves fitting procedures for peak positions and intensities, which can be used to identify the underlying structure. Such analysis of experimental spectra is complicated because it only deals with the statistical average spectrum, and the underlying individual atomistic structures may have considerable variation in their core-level spectra. On the other hand, sometimes appreciable changes in atomic positions cause only subtle spectral changes [3]. Thus, not all structural information is manifested in the spectra of the system, which raises a deeper question of how to find the structural degrees of freedom that the experiment is sensitive to.

To gain more insight into the experiment, calculations may be performed to interpret spectral lines and their connection to structural degrees of freedom. For crystalline materials, existing structural knowledge and geometry optimization can be used to obtain structures for spectral evaluation. This common approach is computationally light and still

allows interpretation of the data in a meaningful fashion [4]. Liquid systems, on the other hand, exhibit a continuous range of configurations. They explore the accessible phase space and, therefore, computational studies of liquids require statistical sampling [5]. This sampling can be accomplished effectively through molecular dynamics simulations, which is the approach we take here.

In this work, we study correlations between structural parameters of the local environment of the absorbing site and intensities in regions of interest (ROI) in the corresponding core-level excitation spectra. Statistical sampling of atomic structures is done by molecular dynamics while spectra are obtained from first-principles calculation. We use liquid water as the benchmark for this procedure, with future studies of systems with a large number of degrees of freedom in mind. Compared to liquid water, ice has stronger tetrahedral structure and hydrogen bonding, and increased intensity in the post-edge region [6] of the O K-edge excitation spectrum. Supercritical water, on the other hand, has more broken hydrogen bonds and greater deviation from the latticelike structure with prominent pre-edge feature in the O K-edge spectrum [7]. The two analysis methods used, classification and linear correlation coefficients, find these structure–spectrum relationships in a single simulation of liquid water.

**II. MODEL AND METHODS**

The nearest neighbors of the excited atomic site (or lack of them) typically cause the most significant effects on its x-ray spectrum. Including the excited atom, we define this atomic cluster as the local configuration  $\mathbf{R}^{(\text{loc})}$ . The thermodynamic distribution causes an ensemble of local configurations, which also yields a distribution of core-level spectra. The challenge is to extract meaningful information by relating aspects of these local structures to intensities of spectral features. First, we define ROI in the simulated ensemble average spectrum, which is followed by integration of intensity from each local configuration over this ROI. This procedure allows us to use a single number,  $F(\mathbf{R}^{(\text{loc})})$ , to quantify the intensity contribution of a local configuration to a particular spectral feature. Second, we define structural parameters, such as the

\*johannes.niskanen@helmholtz-berlin.de

number of hydrogen bonds of the probed atom, as functions of the local configuration. In general, such a parameter is a real-valued quantity  $P(\mathbf{R}^{(\text{loc})})$  for which some mathematical definition needs to be chosen. We therefore relate a local configuration to both its structural parameters and the ROI intensities of the resulting spectrum. While  $F$  and  $P$  do not generally have a 1-to-1 correspondence, statistical correlations between these quantities can be obtained, which give insight into how line intensities are related to underlying structures.

### A. Computational methods

Atomic structural models of water were generated by *ab initio* molecular dynamics (AIMD), as described in detail in Ref. [8]. The system of 64 water molecules was treated in  $(1.24 \text{ nm})^3$  cubic cell with periodic boundary conditions. The system was first equilibrated in the canonical (NVT) ensemble at 330 K for about 20 ps using a velocity rescaling thermostat [9], after which the dynamics was continued in the microcanonical (NVE) ensemble for about 200 ps. The resulting temperature in the NVE simulation was around 340 K. Accounting for the quantum dynamics of the hydrogen nuclei, such as via path integral molecular dynamics (PIMD), gives improved radial distribution functions for water, which then yields spectra with pre-edge features in closer agreement with experiment. PIMD simulations are becoming computationally tractable by using ring-polymer contraction and multiple-time-stepping approaches [10]. However, the cost of spectra calculations still increases linearly with the number of replicas used in the PIMD simulation. The overstructuring of the oxygen-oxygen RDF in our simulations are in part also due the use of an atom-centered, incomplete basis set [11], and hence a corresponding basis-set superposition error [12]. As a compromise, it has been found that increasing the simulation temperature to 330 K in AIMD produces results for the O-O and O-H radial distribution functions in close agreement to 300 K results from PIMD simulations [13]. Even with this *ad hoc* scaling factor, the temperature is close to typical experimental setups, and the systems are not affected by any anomalous properties of water. Also for cosolute solutions, good agreement with experimental results using this setup could be achieved [8].

Forces for the integration of Newton's equation of motion were obtained using density functional theory as implemented in the Quickstep [14] module of CP2K [15]. We used the BLYP exchange-correlation functional [16,17] together with a pairwise-additive dispersion correction [18]. The Kohn-Sham orbitals were expanded into a combined plane-wave and Gaussian orbital basis set, with a kinetic energy cutoff of 400 Ry and a dual- $\zeta$  representation [19], respectively. Core-electrons were replaced by separable dual-space Gaussian pseudopotentials [20]. The O-O radial distribution function from the simulation is shown in Fig. 1(a) in comparison to experiment from Ref. [21]. Due to always-present approximations and limitations, the presented simulation does not provide the exact configurational sampling, seen in the comparison of RDFs. In this case, slight overstructuring of water is seen, but in general the BLYP functional with dispersion correction has been found to give a good description of liquid water [22]. In terms of this work the exact weighting is not necessary as we

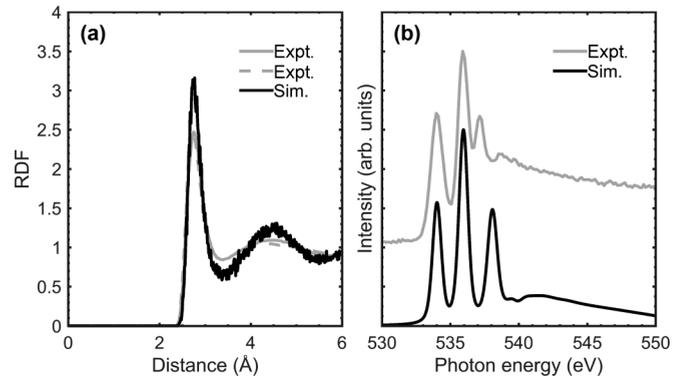


FIG. 1. The O-O radial distribution function from the AIMD simulation compared to experiment (Ref. [21]) (a). The calculated and experimental (Ref. [26]) O1s excitation spectra of gas-phase water (b). The simulated spectrum was shifted to be aligned at the lowest resonance.

are more interested in dependence of ROI intensities on the change of structural parameters.

Individual nonresonant inelastic x-ray scattering spectra were calculated for each oxygen atom of 21 snapshots, sampled with 5-ps intervals, along the production run trajectory by using the Bethe-Salpeter equation (BSE) method as implemented in the OCEAN code [23,24]. It has previously been demonstrated that the BSE method, as implemented in the OCEAN code, can produce satisfactory x-ray absorption spectra of liquid and ice-phase  $\text{H}_2\text{O}$  [25]. For further reference, in Fig. 1(b) we compare the calculated spectrum of a single geometry-optimized  $\text{H}_2\text{O}$  molecule with the experimental gas-phase spectrum taken from Ref. [26]. The sampling procedure of the AIMD simulation yields 1344 spectra, each of which holds unique information about the local chemical and structural environment  $\mathbf{R}_i^{(\text{loc})}$  around the respective excited oxygen atom ( $i = 1, \dots, 1344$ ). Averaging over all individual oxygen K-edge spectra yields the average spectrum (solid black line in Fig. 2), which can be compared to the experimental data (solid gray line in Fig. 2). We used a momentum transfer of  $3.1 \text{ \AA}^{-1}$  for the simulated spectra and convoluted all individual spectra with a GW self-energy obtained from a many-pole approximation to the valence level loss function [27] and a Gaussian of 0.6 eV full width at half maximum (FWHM). These values are typical for experimental non-resonant inelastic x-ray scattering spectroscopy data recorded using the large solid angle spectrometer [28] at the inelastic x-ray scattering beamline ID20 of the European Synchrotron Radiation Facility (ESRF). These are also the values for the presented experimental spectrum of liquid water.

Since OCEAN uses pseudopotentials and yields spectra with respect to the Fermi level of each snapshot, the spectra from different MD simulation snapshots are not necessarily on the same absolute energy scale. To calibrate spectra from snapshot to snapshot, the total energy shift was obtained for the average spectrum of each snapshot such that the average snapshot shows the prepeak at 535 eV. This rigid shift was then applied to all 64 individual spectra from the corresponding snapshot. The simulated gas-phase spectrum of Fig. 1(b) was likewise aligned to the first resonance of the experiment.

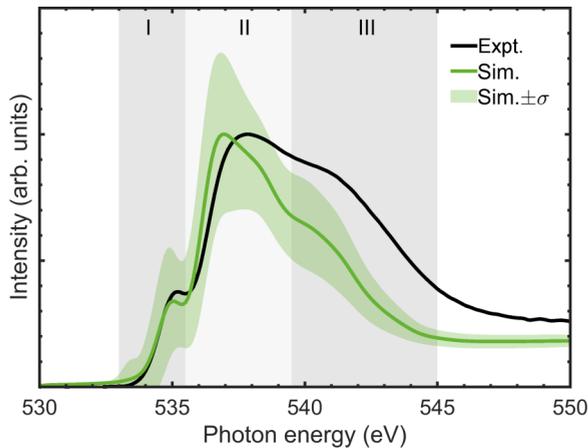


FIG. 2. The simulated ensemble average (green) and the experiment (black). The limits within one standard deviation among the 1344 individual spectra is shown as green shading. The energy regions of interest I–III are shown as gray shading. This ROI selection is based on identifying the pre-edge, main edge, and post-edge in the simulated ensemble average.

### B. Regions of interest

As can be seen from Fig. 1(b), the x-ray spectrum of isolated H<sub>2</sub>O consists of three main features that can be attributed to transitions to anti-bonding molecular orbitals of  $4a_1$ ,  $2b_2$ , and  $3b_2$  symmetry, in order of increasing energy. The symmetries of these orbitals are largely maintained in the condensed phase [29], but are referred to less specifically as the pre-edge, main-edge, and post-edge regions. We use these established features to naturally define three regions of interest (ROIs), specifically I = [533, 535.5], II = [535.5, 539.5], III = [539.5, 545] in eV. We assign a single numeric value for each region of interest to each absorption site by integrating the normalized spectral intensity over the energy range of the region. We label these ROI intensities as  $F_X$  for  $X = I, II, III$ .

### C. Structural characterization

Our objective is to identify correlations between certain structural motifs and changes in the intensities of each ROI. From the molecular dynamics simulation, we extracted 21 statistically independent structures, each consisting of 64 water molecules. This gives us a sample of 1344 unique oxygen environments. In the remainder of this section we define a set of structural parameters based on chemical intuition or common definitions from literature. The next section quantifies the correlations between the values of these structural parameters and the ROI intensities. The basic assumption of this procedure is that spectral intensities are largely determined by the local atomic structure. Thus, we consider only the configuration of the atoms within the first two solvation shells. The number of molecules in different solvation shells is based on oxygen-oxygen distance. The first solvation shell of the central oxygen consists of water molecules with the oxygen atom closer than 3.4 Å to the central oxygen. The second solvation shell consists of molecules with the oxygen atom in the range from 3.4 Å to 5.5 Å from the central oxygen. These radial cutoffs are

based on the minima found in the radial distribution function of Fig. 1(a) after smoothing.

#### 1. Deviation from tetrahedrality

The angular deviation from tetrahedrality of the first solvation shell was defined based on the summed angular deviation from an angle of  $109.5^\circ$ . We define  $\mathbf{r}_i^{\text{loc}}$  as the positions of the four nearest neighbor oxygen atoms ( $i = 1, 2, 3, 4$ ), measured from the excited oxygen. The angular deviation from tetrahedrality,  $\Delta_a$ , was defined and evaluated as

$$\Delta_a = \sum_{i=1}^3 \sum_{j=i+1}^4 |\text{acos}(\hat{\mathbf{r}}_i^{\text{loc}} \cdot \hat{\mathbf{r}}_j^{\text{loc}}) - 109.5^\circ|,$$

where  $\hat{\mathbf{r}}_k^{\text{loc}} = \mathbf{r}_k^{\text{loc}} / |\mathbf{r}_k^{\text{loc}}|$ . Our definition for distance deviation from tetrahedrality,  $\Delta_d$ , evaluates the deviation of furthest and the closest of the four closest neighbors,

$$\Delta_d = \max \{ |\mathbf{r}_i^{\text{loc}}| \}_i - \min \{ |\mathbf{r}_i^{\text{loc}}| \}_i.$$

Parameters  $\Delta_a$  and  $\Delta_d$  are both nonnegative, and apart from rounding errors (in  $109.5$ ) equal to zero for a perfect tetrahedron of any size.

#### 2. Hydrogen bonds

Hydrogen bonding was evaluated for both accepted and donated bonds separately, using the criteria described in Refs. [29,30]. To classify as hydrogen-bonded within the first coordination shell, the oxygen atoms of the two water molecules needed to be separated by less than 3.5 Å and the hydrogen-donor-acceptor angle needed to be less than or equal to  $30^\circ$ . For the absorbing oxygen site, we quantify the number of donating hydrogen bonds (#D), the number of accepting hydrogen bonds (#A), and the total number of hydrogen bonds (#T).

#### 3. Additional Structural Parameters

In addition to the five structural parameters just defined—the number of donated, accepted, or total hydrogen bonds (#D, #A, #T), and parameters measuring the deviation from tetrahedrality for the four nearest water molecules ( $\Delta_a$  and  $\Delta_d$ )—we define six additional parameters. These are the bending angle of the water molecule ( $\phi$ ), the shorter and longer of the OH bonds ( $d_{\text{OH}}^s$  and  $d_{\text{OH}}^l$ ), and the number of water molecules in the first solvation shell (SS1), the second solvation shell (SS2), or in either shell (SS12). The distribution of all evaluated structural parameters and intensities in ROIs I, II, and III are presented in the Appendix.

### D. Correlation analysis

#### 1. Mean-based classification

Classification was performed by first calculating the mean of all intensities in regions I, II, and III. For each interval, each local structure was classified as above-the-average or below-the-average in intensity. The mean of each structural parameter was then evaluated for both above-the-average and below-the-average sets, again for each line region I, II, and III. Finally, the difference between the two was evaluated. Error estimation was based on the 10 000-fold bootstrap algorithm

where the original data set was resampled to data of the size of the original. The analysis procedure was repeated and the standard deviation of the 10 000 results for each value is given as error estimate.

## 2. Linear correlation coefficients

To search for correlated behavior of ROI-intensity and structural parameter we calculated linear correlation coefficients between the two, for each possible ROI-parameter combination. The linear correlation coefficients measure the strength of linear dependence between the ROI intensity and structural parameter value. Therefore, ROI-intensity—structural parameter linear correlation coefficients represent how the intensities in the different ROIs vary with the underlying structural parameters. A positive (negative) sign of the coefficient means an increase (decrease) of the ROI intensity when the parameter value increases. Before the procedure, the mean value was subtracted from each parameter set. Again, error estimation was based on the 10 000-fold bootstrap algorithm where the original data set was resampled to data of the same size. The procedure was repeated and the standard deviation of each value is given as error estimate.

## III. RESULTS

The simulated ensemble averaged oxygen K-edge spectrum of liquid water is presented in Fig. 2, together with the experiment. The channel-wise standard deviation due to statistical variation in the spectrum is depicted as shading. While exact agreement with the experiment is not observed, all the characteristic features of the experiment are identifiably reproduced. This allows linking the results of the simulation to the experiment; the pre-edge, main-edge, and post-edge are clearly resolved, allowing corresponding ROI selection, shown as shaded areas in the background. We assign ROI I as pre-edge, ROI II as main-edge, and ROI III as post-edge.

As an example of the data our procedure bases on, Fig. 3 presents the normalized intensity distribution in the pre-edge (ROI I) using the hydrogen-bonding parameters in Figs. 3(a)–3(c), the deviation-from-tetrahedrality parameters  $\Delta_a$  and  $\Delta_d$  in Figs. 3(d)–3(e), and a generated parameter  $d_{\text{OH}}^s + d_{\text{OH}}^l$  in Fig. 3(f). The structure-intensity data show lots of scatter and the structural dependence is seen only as a relatively weak trend. This implies that typically used structural terminology does not provide categorical explanation for spectral effects, and the underlying dependencies are much more complex. Still, the parameters related to hydrogen bonds show an increase in pre-edge intensity when donated or accepted hydrogen bonds are broken. Also, deviation from a tetrahedral geometry causes similar effects.

We now turn to the quantified intensity—local structure parameter interpretation. We approach the subject using two methods: mean-based classification and linear correlation coefficients. We underline that the procedure is aimed for the general case, and that liquid water is chosen because it provides the best-studied benchmark. Our motivation here is to study two simple methods, which naturally means that the list is not complete.

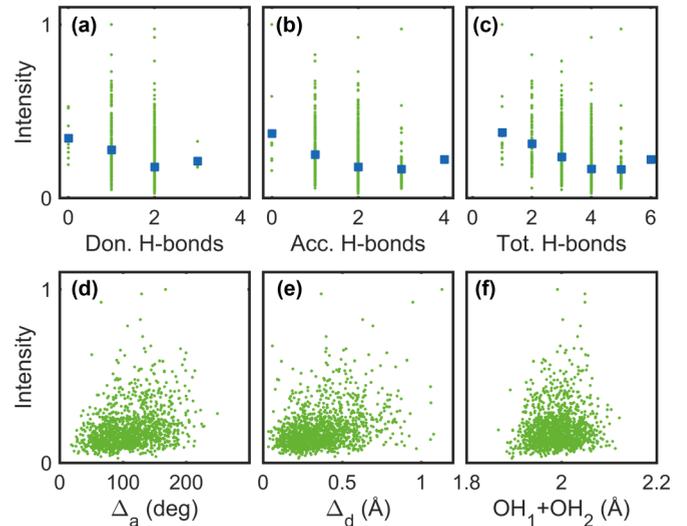


FIG. 3. The integrated pre-edge (ROI I) intensity of the 1344 individual sites depicted as a function of chosen structural parameters: donated, accepted, and total hydrogen bonds (a–c),  $\Delta_a$  and  $\Delta_d$  (d, e), and a generated parameter  $d_{\text{OH}}^s + d_{\text{OH}}^l$  (f). For the hydrogen bonding parameters, the average is shown in blue.

### 1. Mean-based classification

An example of a simple classification of structural parameter-ROI intensity ( $P$ - $F$ ) relation is based on whether a marginal increase of the weight of a local configuration in the statistical distribution increases or decreases the intensity in a given ROI. We performed this classification by first comparing intensities in ROIs I, II, and III for each of the 1344 spectra to the corresponding ROI intensity of the average spectrum, and then identifying which configurations boost or suppress intensity in each ROI. Difference between the average values for the structural parameters of the two sets are presented in Fig. 4 normalized to the mean values. A table of the absolute difference is given in the Appendix. A positive value means that larger parameter values result in an increase of ROI intensity. The error estimates represent standard deviations of the parameters in a bootstrap procedure of 10 000 resamplings of the data.

The data from Fig. 4 reveal how different ROIs are sensitive to distinct structural motifs. For example, we can observe a correlation between the intensity in the pre-edge region (ROI I) and the deviation-from-tetrahedral parameters: the configurations showing above-average intensity in the pre-edge region have higher average parameters  $\Delta_a$  and  $\Delta_d$ , which is manifested as positive corresponding values in Fig. 4. Contrariwise, intensity in the post-edge ROI III decrease with increasing deviation from tetrahedrality. This is in line with the commonly accepted qualitative interpretation scheme of water: tetrahedrally ordered ice phases show prominent post-edge features [6], whereas water in the liquid and especially in the supercritical regime exhibits very prominent pre-edges [7]. Similarly, the increasing numbers of hydrogen bonds show decrease (increase) of the pre-edge (post-edge) in line with data from supercritical water and ice.

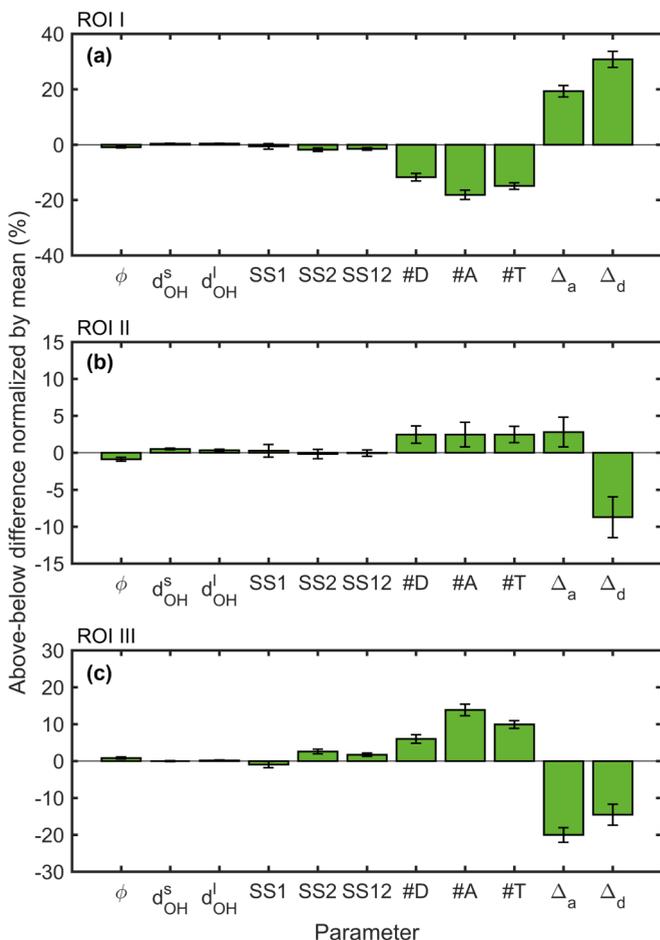


FIG. 4. Difference between mean structural parameters for structures boosting and suppressing average ROI intensity. Positive values mean that an increase in the parameter is associated with an increase in ROI intensity. The absolute values were normalized by the mean.

## 2. Linear correlation coefficients

We calculated the  $P$ - $F$  linear correlation coefficients, which are presented in Fig. 5, and as a table in the Appendix. In each case, the mean value of the data was subtracted before calculating the coefficients and errors represent the standard deviation in a 10 000-fold bootstrap re-sampling. The results presented in Fig. 5 mostly show similar behavior as derived from the classification scheme. The analysis shows that the pre-edge intensity depends positively on broken tetrahedrality  $\Delta_a$ , and  $\Delta_d$ , whereas the opposite is the case for the post edge, ROI III. In a similar manner, the pre-edge intensity is negatively correlated with the number of hydrogen bonds, whereas the opposite is seen for the post edge ROI III.

## IV. DISCUSSION

The oxygen K-edge spectrum of water has been studied intensively during the past 15 years [6,7,29,31–34] and reviewed recently [35]. Although the structure of water is still under debate [34], certain spectral-structure relations have gradually become apparent and a common consensus seems to be emerging about features in the oxygen K-edge spectrum of liquid water. The pre-edge is enhanced together with the

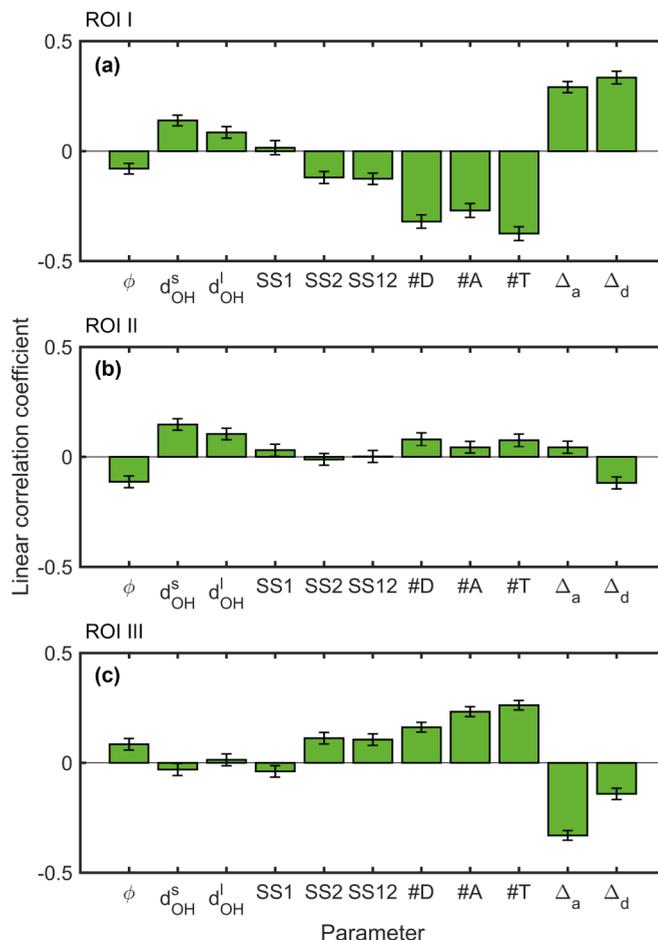


FIG. 5. Linear correlation coefficients between structural parameters and intensities in ROIs I, II, and III.

main edge in supercritical water [7] and therefore is associated with broken hydrogen bonds. The post-edge, on the other hand, is enhanced in ice [6] and therefore indicates strong tetrahedral order and hydrogen bonding. The analysis methods presented herein succeeded in finding these spectrum-property correlations.

The pre-edge (ROI I) is prominent in supercritical water [7] and it gains spectral weight and shifts to lower energies for increasing temperature in the liquid state [33]. It has, therefore, been related to weakened or broken hydrogen bonds and a disturbed hydrogen bond network. However, the pre-edge also changes as a function of the approaching second solvation shell in high-pressure ice phases [32], which indicates the complexity of the interrelation and spectrum-structure relations. The main-edge (ROI II) is also commonly connected with a heavily distorted and weakly hydrogen bonded water network and, for example, Tse *et al.* reported a decrease in spectral weight in the main-edge when comparing high- and low-density amorphous water ice [6]. As for the case of the pre-edge feature, changes in the main-edge have also been attributed to structural changes in which the hydrogen bonding remains the same; i.e., Pylkkänen *et al.* reported a linear increase of the main-to-post-edge intensity ratio when comparing different ice phases of different densities but similar hydrogen bond arrangements [32]. The post-edge (ROI III), in

contrast, is prominent in presence of tetrahedral order and a strong hydrogen-bond network. Thus, the oxygen K-edges of several ice phases exhibit increased spectral weight in the post-edge region [6,32]. Tetrahedrality and hydrogen bonding are linked in our analysis to post-edge intensity.

It is fascinating that by studying a liquid system in only one statistical ensemble, it is possible to obtain structure-spectrum (ROI intensities) correlations seen in experiments for both ice and supercritical water. This is owing to the *variation* of local configurations in the ensemble. Thus, the analysis procedure outlined can be expected to work for cases where  $\rho$  changes due to a change in experimental conditions, but the system is not pushed to completely new parts of the local phase space. For example, our liquid simulation does not probe high-density-ice-like local configurations well, because these involve reorganization of the whole second solvation shell that is very unlikely to happen in the liquid state. As a result, the main edge behavior seen in Ref. [32] is not probed by our analysis.

Water has been proposed to appear in two coexisting liquid phases [34]. Some MD simulations have shown these two phases [36], but these results have also been questioned [37]. Studying this problem with our method would be very complicated. First, real-valued mathematical definitions for the phases are required, and the results would only consider the particular definition out of its infinitely many alternatives. Second, the structural simulations should not limit two-phase phenomena (if they exist) simply because of a limited box size and time scale. Moreover, the spectral simulation should be able to treat such boxes. All this is beyond 64-molecule systems, and therefore we limit the scope of this work to simple, maximally local, and well-understood structural parameters, for which a general consensus in terms of spectral changes along structure has been established. We note that there is indeed a wealth of liquid systems for which phenomena related to parameters like these are the main interest of the simulation.

For the applied method, the nature of the spectroscopic transition is not very relevant as long as the process defines a spectrum depending dominantly on the local configuration. Due to enormous amounts of transition lines even for one local configuration, core-level absorption, emission, and resonant inelastic scattering yield  $\mathcal{R}^m$ -valued functions, where  $m$  in practice is the number of energy channels in the relevant energy range. For a single local configuration, on the other hand, K-edge x-ray photoelectron spectroscopy typically yields a real number (the core-electron binding energy). From the point of view of this work, the more complicated spectral signals like x-ray absorption may therefore provide more information about the sample.

## V. CONCLUSIONS

We have presented a systematic statistical analysis of simulations to help the interpretation of core-level spectra. Our approach defines regions of interest in the spectra, based on identifiable spectral features in a simulated ensemble average. The intensity of these regions allows for a quantitative analysis and interpretation of the spectral features in terms of the underlying local structural parameters. For our analysis we

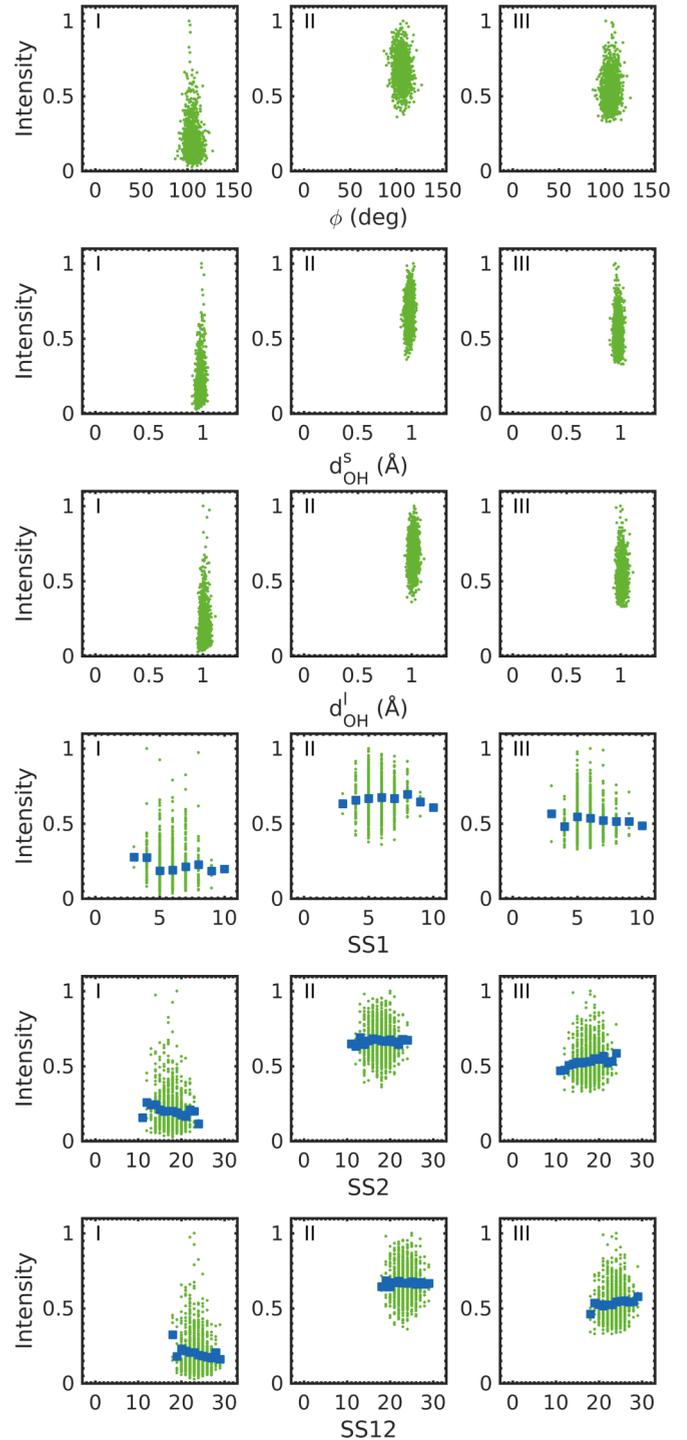


FIG. 6. The intensity data in regions I, II, and III for bonding angle, for the shorter OH bond, for the longer OH bond, and for numbers of water molecules in solvation shells. Each green marker corresponds to a local configuration in the simulation. The blue markers show the mean intensity value for the discrete-valued parameter values.

studied a simulation of the oxygen K-edge spectrum of liquid water, and while our simulation does not achieve a perfect match with the experiment, all observed spectral features were reproduced identifiably. Based on extensive sampling we find that statistical distribution of spectral intensities of these

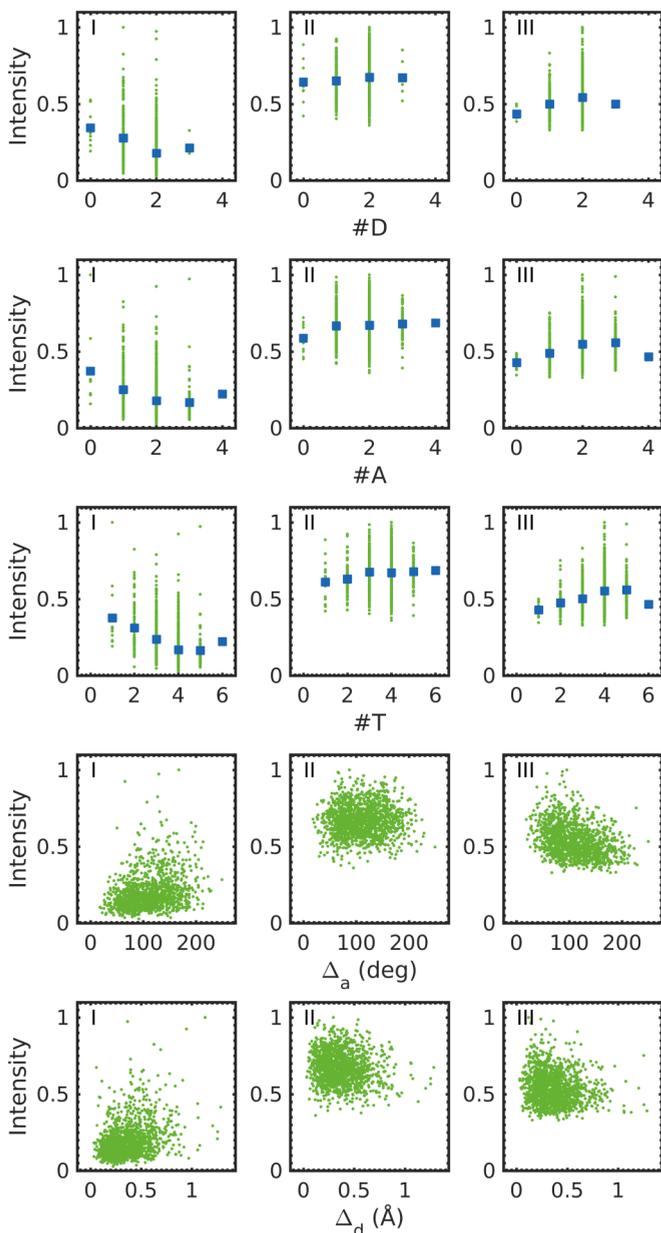


FIG. 7. The intensity data in regions I, II, and III for numbers of hydrogen bonds made by the molecule, and for deviation from the tetrahedral geometry. Each mark corresponds to a local configuration in the simulation. The blue markers show the mean intensity value for the discrete-valued parameter values.

features as a function of structural parameters is heavily scattered, but trends from the known experimental data of liquid water, supercritical water, and ice are still underlying and are found by the analysis procedure. Our results show that due to statistical variation, core-level spectra can also be interpreted by using MD simulations and by studying correlations between spectral intensities and structural parameters, instead of fingerprint analysis using predefined model structures.

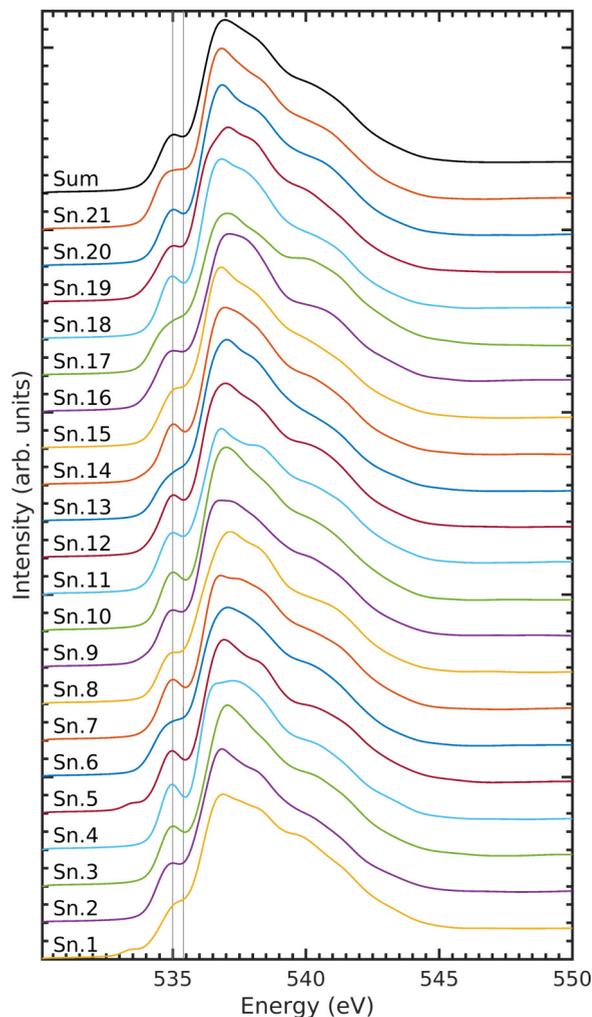


FIG. 8. Alignment of the 64-oxygen spectra from the different snapshots. The vertical lines at 535.0 and 535.4 eV are presented to guide the eye.

#### ACKNOWLEDGMENTS

We are grateful to Dr. I. Juurinen and to Dr. A. Kettunen for related discussions. We acknowledge the ESRF for providing synchrotron radiation. C. Henriquet is kindly thanked for technical support. C.J.S. is thankful to M. Moretti-Sala for supporting this work.

#### APPENDIX

Figures 6 and 7 present intensities in regions I, II, and III against the structural parameters. Each point corresponds to a local configuration in the simulation. Figure 8 represents the alignment of the spectra for the 64-spectrum snapshots. Table I presents the parameter values for the classification scheme and Table II presents the linear correlation coefficients.

TABLE I. Difference between mean structural parameters for structures boosting and suppressing average ROI intensity. Positive values mean that an increase in the parameter is associated with an increase in ROI intensity. The total average is also given.

Parameter	I	II	III	Ave.
$\phi$	-0.93(0.29)	-0.92(0.28)	0.87 (0.28)	105.87
$d_{\text{OH}}^s$	0.00(0.00)	0.01(0.00)	0.00(0.00)	0.98
$d_{\text{OH}}^l$	0.00(0.00)	0.00(0.00)	0.00(0.00)	1.01
SS1	-0.03(0.06)	0.02(0.05)	-0.05(0.05)	5.70
SS2	-0.32(0.12)	-0.03(0.11)	0.46(0.11)	17.86
SS12	-0.36(0.10)	-0.01(0.10)	0.41(0.10)	23.56
#don	-0.21(0.03)	0.05(0.02)	0.11(0.02)	1.84
#acc	-0.33(0.03)	0.05(0.03)	0.25(0.03)	1.84
#tot	-0.55(0.04)	0.09(0.04)	0.36(0.04)	3.67
$\Delta_a$	21.29(2.25)	3.10(2.24)	-22.06(2.15)	110.16
$\Delta_d$	0.11(0.01)	-0.03(0.01)	-0.05(0.01)	0.37

TABLE II. Linear correlation coefficients between structural parameters and intensities in ROIs I, II, and III.

Parameter	I	II	III
$\phi$	-0.08(0.02)	-0.11(0.03)	0.08(0.03)
$d_{\text{OH}}^s$	0.14(0.02)	0.15(0.03)	-0.03(0.03)
$d_{\text{OH}}^l$	0.09(0.03)	0.10(0.03)	0.01(0.03)
SS1	0.02(0.03)	0.03(0.03)	-0.04(0.03)
SS2	-0.12(0.03)	-0.01(0.03)	0.11(0.03)
SS12	-0.13(0.03)	0.00(0.03)	0.11(0.03)
#don	-0.32(0.03)	0.08(0.03)	0.16(0.02)
#acc	-0.27(0.03)	0.04(0.03)	0.23(0.02)
#tot	-0.38(0.03)	0.08(0.03)	0.26(0.02)
$\Delta_a$	0.29(0.03)	0.04(0.03)	-0.33(0.02)
$\Delta_d$	0.33(0.03)	-0.12(0.03)	-0.14(0.03)

- [1] M. Tuckerman, *Statistical Mechanics. Theory and Molecular Simulation* (Oxford University Press, Oxford, 2010).
- [2] J. Niskanen, N. A. Murugan, Z. Rinkevicius, O. Vahtras, C. Li, S. Monti, V. Caravatta, and H. Ågren, *Phys. Chem. Chem. Phys.* **15**, 244 (2013).
- [3] J. Niskanen, C. J. Sahle, I. Juurinen, J. Koskelo, S. Lehtola, R. Verbeni, H. Müller, M. Hakala, and S. Huotari, *J. Phys. Chem. B* **119**, 11732 (2015).
- [4] J. Inkinen, J. Niskanen, T. Talka, C. Sahle, H. Müller, L. Khriachchev, J. Hashemi, A. Akbari, M. Hakala, and S. Huotari, *Sci. Rep.* **5**, 15851 (2015).
- [5] M. Allen and D. Tildesley, *Computer Simulation of Liquids* (Oxford Science Publications, Oxford, 1987).
- [6] J. S. Tse, D. M. Shaw, D. D. Klug, S. Patchkovskii, G. Vankó, G. Monaco, and M. Krisch, *Phys. Rev. Lett.* **100**, 095502 (2008).
- [7] C. J. Sahle, C. Sternemann, C. Schmidt, S. Lehtola, S. Jahn, L. Simonelli, S. Huotari, M. Hakala, T. Pylkkänen, A. Nyrow, K. Mende, M. Tolan, K. Hämäläinen, and M. Wilke, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6301 (2013).
- [8] M. B. Hahn, F. Uhlig, T. Solomun, J. Smiatek, and H. Sturm, *Phys. Chem. Chem. Phys.* **18**, 28398 (2016).
- [9] G. Bussi, D. Donadio, and M. Parrinello, *J. Chem. Phys.* **126**, 014101 (2007).
- [10] O. Marsalek and T. E. Markland, *J. Chem. Phys.* **144**, 054112 (2016).
- [11] H.-S. Lee and M. E. Tuckerman, *J. Chem. Phys.* **125**, 154507 (2006).
- [12] M. V. Houteghem, T. Verstraelen, A. Ghysels, L. Vanduyfhuys, M. Waroquier, and V. V. Speybroeck, *J. Chem. Phys.* **137**, 104506 (2012).
- [13] J. A. Morrone and R. Car, *Phys. Rev. Lett.* **101**, 017801 (2008).
- [14] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, *Comput. Phys. Commun.* **167**, 103 (2005).
- [15] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele, *Wiley Interdisc. Rev.: Comput. Mol. Sci.* **4**, 15 (2014).
- [16] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- [17] C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785 (1988).
- [18] S. Grimme, *J. Comput. Chem.* **27**, 1787 (2006).
- [19] J. VandeVondele and J. Hutter, *J. Chem. Phys.* **127**, 114105 (2007).
- [20] S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* **54**, 1703 (1996).
- [21] A. Soper, *J. Phys.: Condens. Matter* **19**, 335206 (2007).
- [22] M. J. Gillan, D. Alf, and A. Michaelides, *J. Chem. Phys.* **144**, 130901 (2016).
- [23] J. Vinson, J. J. Rehr, J. J. Kas, and E. L. Shirley, *Phys. Rev. B* **83**, 115106 (2011).
- [24] K. Gilmore, J. Vinson, E. Shirley, D. Prendergast, C. Pemmaraju, J. Kas, F. Vila, and J. Rehr, *Comput. Phys. Commun.* **197**, 109 (2015).
- [25] J. Vinson, J. J. Kas, F. D. Vila, J. J. Rehr, and E. L. Shirley, *Phys. Rev. B* **85**, 045101 (2012).
- [26] I. Ishii, R. McLaren, A. P. Hitchcock, and M. B. Robin, *J. Chem. Phys.* **87**, 4344 (1987).
- [27] J. J. Kas, A. P. Sorini, M. P. Prange, L. W. Cambell, J. A. Soininen, and J. J. Rehr, *Phys. Rev. B* **76**, 195116 (2007).
- [28] S. Huotari, C. J. Sahle, C. Henriquet, A. Al-Zein, K. Martel, L. Simonelli, R. Verbeni, H. Gonzalez, M.-C. Lagier, C. Ponchut, M. Moretti Sala, M. Krisch, and G. Monaco, *J. Synchrotron Radiat.* **24**, 521 (2017).
- [29] W. Chen, X. Wu, and R. Car, *Phys. Rev. Lett.* **105**, 017802 (2010).
- [30] A. Luzar and D. Chandler, *Phys. Rev. Lett.* **76**, 928 (1996).
- [31] P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odellius, H. Ogasawara, L. Å. Näslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson, *Science* **304**, 995 (2004).
- [32] T. Pylkkänen, V. M. Giordano, J.-C. Chervin, A. Sakko, M. Hakala, J. A. Soininen, K. Hämäläinen, G. Monaco, and S. Huotari, *J. Phys. Chem. B* **114**, 3804 (2010).
- [33] T. Pylkkänen, A. Sakko, M. Hakala, K. Hämäläinen, G. Monaco, and S. Huotari, *J. Phys. Chem. B* **115**, 14544 (2011).
- [34] A. Nilsson and L. G. M. Pettersson, *Nature Commun.* **6**, 8998 (2015), and references therein.
- [35] T. Fransson, Y. Harada, N. Kosugi, N. A. Besley, B. Winter, J. J. Rehr, L. G. M. Pettersson, and A. Nilsson, *Chem. Rev.* **116**, 7551 (2016).
- [36] T. Yagasaki, M. Matsumoto, and H. Tanaka, *Phys. Rev. E* **89**, 020301 (2014).
- [37] S. D. Overduin and G. N. Patey, *J. Chem. Phys.* **143**, 094504 (2015).