An Exact *A Posteriori* Correction for Hydrogen Bond Population Correlation Functions and Other Reversible Geminate Recombinations Obtained from Simulations with Periodic Boundary Conditions. Liquid Water as a Test Case

Johanna Busch,[1] Jan Neumann,[1] and Dietmar Paschek[1, *]

[1] *Institut für Chemie, Abteilung Physikalische und Theoretische Chemie, Universität Rostock, Albert-Einstein-Str. 21, D-18059 Rostock, Germany*

(Dated: 06/01/2021; DOI: 10.1063/5.0053445)

The kinetics of breaking and re-formation of hydrogen bonds (HBs) in liquid water is a prototype of a reversible geminate recombination. HB population correlation functions (HBPCFs) are a means to study HB kinetics. The long-time limiting behaviour of HBPCFs is controlled by translatoric diffusion and shows a $t^{-3/2}$ time-dependence, which can be described by analytical expressions based on the HB acceptor density and the donor-acceptor inter-diffusion coefficient. If the trajectories are not properly "unwrapped", the presence of periodic boundary conditions (PBCs) can perturb this long-time limiting behaviour. Keeping the trajectories "wrapped", however, allows for a more efficient calculation of the HBPCFs. We discuss the consequences of PBCs in combination with "wrapped" trajectories following from the approximations according to Luzar-Chandler and according to Starr, each deviating in a different fashion from the true long-time limiting behaviour, but enveloping the unperturbed function. A simple expression is given for estimating the maximum time up to which the computed HBPCFs reliably describe the long-time limiting behaviour. In addition, an exact *a posteriori* correction for systems with periodic boundary conditions for "wrapped trajectories" is derived, which can be easily computed and which is able to fully recover the true $t^{-3/2}$ long-time behaviour. For comparison, HBPCFs are computed from MD simulations of TIP4P/2005 model water for varying system sizes and temperatures of 273 K and 298 K using this newly introduced correction. Implications for the computations of HB lifetimes and the effect of the system-size are discussed.

I. INTRODUCTION

The prototypical hydrogen-bonded liquid is, of course, water.[1] According to Frank Stillinger, *the key to understanding liquid water and its solutions lies in the concept of the "hydrogen bond"*.[2] As a consequence, the understanding of the hydrogen bond (HB) formation processes provides critical insight into the dynamical properties of the liquid. Hydrogen bonds, however, have been shown to be crucial for the behaviour of many other molecular and ionic liquids [3, 4], and are essential for protein stability and many biophysical processes.[5] The code for life itself is written and copied based on matching hydrogen bonding patterns.[6]

Regarding HB dynamics in water, Luzar and Chandler [7] have argued that the HB dynamics are characterised by local relaxation processes, which are mostly uncorrelated with respect to the specific bonding patterns near a certain hydrogen bond.[8, 9] They could show that diffusion processes govern whether one particular pair of water molecules are close neighbours, and HBs between such pairs form and persist randomly with mean lifetimes determined by rate constants for bond making and breaking.

A particularly popular way to study HB dynamics is the concept of "intermittent" HB lifetimes, determined via integrals over hydrogen bond population correlation functions (HBPCFs) as introduced by Stillinger, Rapaport and others.[10–17] From the point of view of chemical kinetics, the breaking and re-formation of a HB can be understood as a reversible diffusion-influenced geminate recombination.[18] The long-time limiting behaviour of HBPCFs, which is well extending into the nanosecond regime for water between 273 K and 298 K, is entirely controlled by translatoric diffusion of the water molecules. For sufficiently long times HBPCFs are approaching the theoretically predicted $t^{-3/2}$ scaling law.[19] This behaviour can be quantitatively described by analytical expressions based on the HB acceptor density and the donor-acceptor inter-diffusion coefficients. The short-time decay, on the other hand, is related to kinetics of HB breaking [8] and the local dynamics of the OH bond-vector, including librations and reorientational motions.[20, 21] The presence of periodic boundary conditions, however, perturbs the long-time limiting behaviour, leading to a systematic under- or overestimation, depending on which particular long-time treatment/approximation for the HBPCFs is used. According to Markovitch and Agmon the occurrence of this perturbation can be avoided, if HB bonded partners are treated as individuals, distinct from their periodic images.[18, 22] These effects are, however, often neglected, with the hope that recombination events with periodic images occur sufficiently rarely.[23, 24] In this communication, we analyze the effect of PBCs on such "wrapped" trajectories by employing random walkers. Here the term *"wrapped" trajectories* refers to the time evolution of molecule positions always "wrapped" into the central box, which is

* dietmar.paschek@uni-rostock.de

often the default mode of how trajectory data are stored. We systematically vary the box-size of a periodic unit-cell to determine the time interval up to which a reliable long-time limiting behaviour can be obtained. The relevant parameters controlling this time interval are identified as the box volume and the translatoric inter-diffusion coefficient of hydrogen bond donors and acceptors. Finally, an exact time-dependent *a posteriori* correction for systems with periodic boundary conditions is derived. This term is able to fully restore the true $t^{-3/2}$ long-time limiting behaviour for $C(t)$ obtained from untreated "wrapped" trajectories, and is just based on the knowledge of the donor-acceptor inter-diffusion coefficient and the box-dimensions. These relations and corrections are tested against MD simulations of liquid TIP4P/2005 water for varying system sizes and temperatures.

II. METHODS

A. MD Simulations

We have performed MD simulations of liquid water using the TIP4P/2005 model [25], which has been demonstrated to rather accurately describe the properties of water compared to other simple rigid nonpolarizable water models.[26] The simulations are carried out at 273 K and 298 K under NVT conditions using system-sizes of 128, 256, and 1024 molecules. MD simulations of 10 ns length each were performed using GROMACS 5.0.6.[27, 28] The integration time step for all simulations was 2 fs. The temperature of the simulated systems was controlled employing the Nosé-Hoover thermostat [29, 30] with a coupling time $\tau_T = 1.0$ ps. Both, the Lennard-Jones and electrostatic interactions were treated by smooth particle mesh Ewald summation.[31–33] The Ewald convergence parameter was set to a relative accuracy of the Ewald sum of 10^{-5} for the Coulomb- and 10^{-3} for the LJ-interaction. All bond lengths were kept fixed during the simulation run and distance constraints were solved by means of the SETTLE procedure.[34]

To determine hydrogen bond population correlation functions, autocorrelation functions over relatively large time sets have to be computed. To evaluate time correlation functions for large time sets with up to 5×10^4 entries efficiently, we applied the convolution theorem using fast Fourier transformations (FFT).[35, 36] The computation of the properties from MD simulations were done using home-built software based on the MDANALYSIS [37, 38], NUMPY [39], and SCIPY [40] frameworks.

B. Random Walkers

Here we outline our use of a "generic" random walker, exploring space with and without periodic boundary conditions. As illustrated in Figure 1, a "hydrogen-bonded" state is defined by the random walker being located within a spherical volume of radius $R_s = 1$ around the center of a cubic box of varying size b with a minimum box-length of $b = 2R_s = 2$. Each walker starts from the origin at $t = 0$ in a hydrogen-bonded state ($h(0) = 1$). New coordinates are computed for discrete time intervals $\delta t = 0.1$ from $\vec{r}(t+\delta t) = \vec{r}(t) + \vec{d}$, where $\vec{d}$ is vector with random orientation and $|\vec{d}| = (6D_{\text{jump}}\delta t)^{1/2}$. Periodic boundary conditions are applied in the sense that the diffusing particle, when leaving the box on one side, will enter on the opposite side. This feature, however, can also be turned off. The HBPCFs reported here are computed by sampling over 10^7 trajectories.

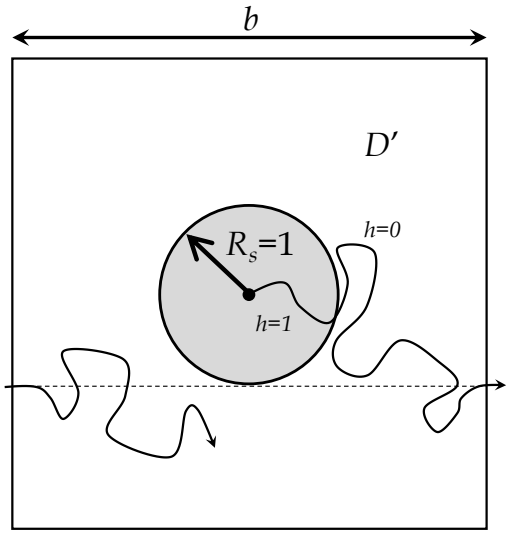

Figure 1. Schematic representation of the "generic" random walker travelling with a diffusion coefficient $D_{\text{jump}} = D'$ with and without ($b \to \infty$) periodic boundary conditions. The spherical volume with radius $R_s = 1$ located at the center of the box defines the "hydrogen bonded" state.

C. "Intermittent" Hydrogen Bond Lifetimes and Dynamics

A popular way to study HB dynamics is the concept of "intermittent" HB lifetimes, introduced by Stillinger, Rapaport and others.[10–17] Here, we adopted the pairwise donor-acceptor definition according to Luzar and Chandler.[8] Therefore, we calculate intermolecular HB population functions $h(t)$ for all possible donor-acceptor pairs in our system. A donor-acceptor pair consists of the OH bond-vector of one water molecule and the oxygen atom of another. A "flipping" of an existing HB, where the former donor becomes the acceptor of a HB from the former acceptor, is considered a different donor-acceptor pair.

The fluctuations of the HB population are described

by the HB population correlation function [9, 11]

$$C(t) = \frac{\langle h(0)h(t)\rangle - \langle h\rangle^2}{\langle h^2\rangle - \langle h\rangle^2} \tag{1}$$

with the HB population function

$$h(t) = \begin{cases} 1, \text{if HB exists} \\ 0, \text{if no HB exists} \end{cases} , \tag{2}$$

which is calculated for every donor-acceptor pair separately. The brackets denote averaging over all donor-acceptor pairs of a specific HB species and all times $t = 0$. The "intermittent" correlation function $C(t)$ describes the fraction of HBs still intact at time t, provided it was intact at $t_0 = 0$, without the need for it to be intact over the whole time interval $t - t_0$. Due to the property of $h(t)$ with $h(t) = h^2(t)$ and thus $\langle h\rangle = \langle h^2\rangle$, the denominator in Equation 1 can be expressed as $[\langle h^2\rangle - \langle h\rangle^2] = \langle h\rangle\,[1 - \langle h\rangle]$ with $\langle h\rangle\,[1 - \langle h\rangle] \approx \langle h\rangle$ for $\langle h\rangle \ll 1$. According to Luzar and Chandler, assuming $\langle h\rangle^2 \approx 0$ is leading to:

$$C(t) \approx \frac{\langle h(0)h(t)\rangle}{\langle h\rangle} . \tag{3}$$

As pointed out by Starr et al.[16], the approximation $\langle h\rangle^2 \approx 0$ in the numerator of Equation 1 could lead to a misrepresentation of $C(t)$ at long time intervals t for "wrapped trajectories" of finite systems with periodic boundary conditions. In particular, $C(t)$ according to Equation 3 does not decay to zero, which poses a problem if you want to compute HB lifetimes by integrating over $C(t)$. Therefore, due to his defence of Equation 1, we will refer to Equation 1 in the remainder of this paper as the "Starr" approximation. Note that for $\langle h\rangle \ll 1$, which is typically well fulfilled for molecular simulations of $N > 100$, Equation 1 is "practically identical" with

$$C(t) \approx \frac{\langle h(0)h(t)\rangle}{\langle h\rangle} - \langle h\rangle . \tag{4}$$

For macroscopic systems, of course, the contribution due to "$-\langle h\rangle$" will be vanishingly small. We will show, however, that for finite sizes and the presence of periodic boundary conditions both, the "Luzar-Chandler"- and "Starr"–approximations envelop the true (i.e. no PBCs) long-time limiting behaviour of $C(t)$. They can only be trusted up to a certain time interval $t_{\max}$ and need to be corrected in a more appropriate fashion.

III. RESULTS AND DISCUSSION

A. Behaviour of a "Generic" Random Walker with Periodic Boundary Conditions

It has been demonstrated, that the long-time behaviour of the hydrogen bond population correlation function in an unrestricted three dimensional environment exhibits a $t^{-3/2}$ time-dependence [41–44], which also follows directly from solving Fick's equation for an instantaneous point source assuming an isotropic and homogeneous diffusion coefficient [45]. For particles hopping between discrete bonded states, the long-time behaviour can be quantitatively expressed by

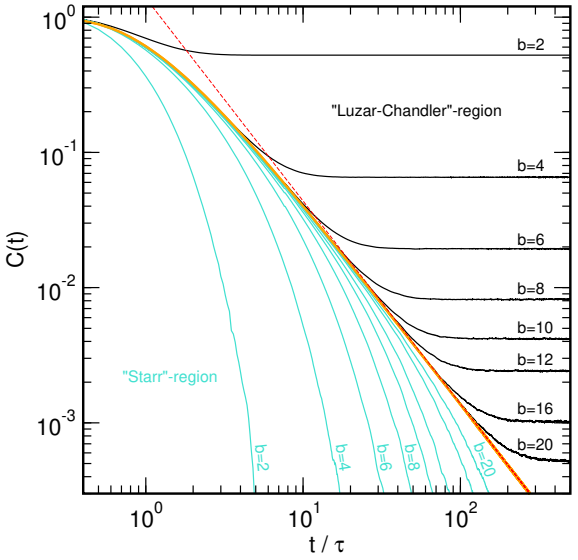

Figure 2. Hydrogen bond population correlation functions $C(t)$ for the "generic" random walker model with $D_{\text{jump}} = 1/6$. Solid black lines: HBPCFs for a random walker with periodic boundary conditions with cubic box-lengths b according to Equation 3. Solid turqoise lines: HBPCFs for a random walker with periodic boundary conditions with the cubic box-lengths b according to Equation 1. In all cases b is given in units of R_s. Solid orange line: Analytical representation for $C_{\text{no-pbc}}(t)$ according to Equation 6. Dashed red line: Analytical representation for $\lim_{t\to\infty} C(t)$ according to Equation 5.

$$\lim_{t\to\infty} C(t) = \frac{1}{s\,\rho_{\text{acc}}(4\pi D'\,t)^{3/2}} , \tag{5}$$

where D' is the inter-diffusion coefficient $D' = D_{\text{donor}} + D_{\text{acceptor}}$ of hydrogen bond donors and acceptors, ρ_{acc} is the HB acceptor-site density, and s is a scaling parameter, depending on the topology of the "network" of hydrogen bond accepting sites, formed by the HB switching pathways interconnecting adjacent sites. For instance, a random walker on a primitive cubic lattice with six next neighbours is described by $s = 1/2$ at even time-steps (see section III B for further clarification). Note that the parameter $(s\,\rho_{\text{acc}})^{-1}$ is essentially assigning a volume to the initial hydrogen-bonded state $h(0)$. To describe the long-time behaviour of the "generic" random walker model, we use here $(s\,\rho_{\text{acc}})^{-1} = 4/3\pi R_s^3$ and $D' = D_{\text{jump}}$.

A quantitative description of $C(t)$ for the "generic" walker model in absence of any periodic boundary conditions follows from the spherical volume defining a hydrogen-bonded state, characterised by a radius R_s.

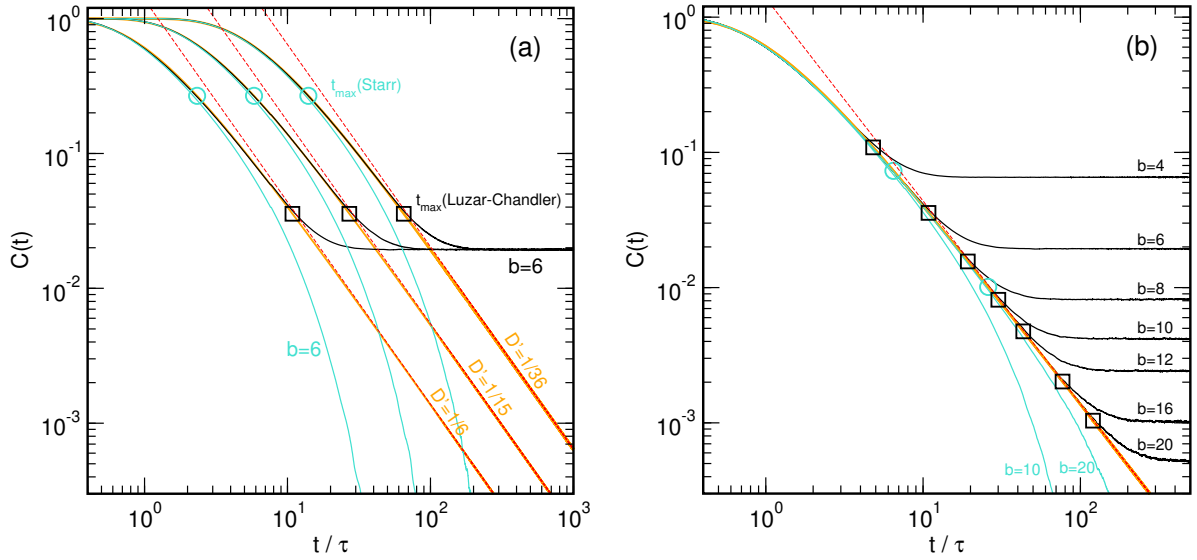

Figure 3. More hydrogen bond population correlation functions $C(t)$ for the "generic" random walker model. Solid black lines: HBPCFs for a random walker with periodic boundary conditions and a box-size b according to Equation 3. Solid turqoise lines: HBPCFs for a random walker with periodic boundary conditions and a box-size b according to Equation 1. Solid orange lines: Analytical representation for $C_{\text{no-pbc}}(t)$ according to Equation 6. Dashed red lines: Analytical solution for $\lim_{t\to\infty} C(t)$ according to Equation 5. The open symbols indicate the time-limits $t_{\max}$ according to Equation 8 up to which the computed HBPCFs reliably describe the long-time limiting behaviour for the "Luzar-Chandler" approximation with $n=2$ and the "Starr" correction with $n=20$. a) Varying Diffusion coefficients with $D' = D_{\text{jump}} = 1/6{,}1/15$, and $1/36$ for a box-size $b = 6$ b) Varying box-sizes b for a diffusion coefficient $D' = D_{\text{jump}} = 1/6$. In all cases b is given in units of R_s.

By integrating the solution of the Fick equation for an instantaneous point source [45] over a spherical volume with radius R_s around the starting point at $t = 0$, the following analytical expression is obtained:

$$C_{\text{no-pbc}}(t) = \text{erf}\left[\frac{R_s}{\sqrt{4D't}}\right] - \frac{R_s}{\sqrt{\pi D't}} \exp\left[-\frac{R_s^2}{4D't}\right], \quad (6)$$

here $D' = D_{\text{donor}} + D_{\text{acceptor}} = D_{\text{jump}}$. For $t \to \infty$, Equation 6 approaches asymptotically Equation 5, finally also exhibiting $t^{-3/2}$ behaviour [42].

The presence of periodic boundary conditions (PBCs) leads to the fact that the initially hydrogen-bonded partners cannot fully escape from one another, since reincarnations in the form of periodic images will reconvene repeatedly. This introduces a deviation for long times t from the $t^{-3/2}$ long-time limiting behaviour and the emergence of a plateau. The plateau value C_∞ is representing the portion of time the hydrogen-bonded partners spend with each other in equilibrium. For the "generic" walker this value corresponds exactly to the ratio of the volume used for defining the HB state and the box volume

$$C_\infty = \frac{4}{3}\pi \frac{R_s^3}{\langle V \rangle} \quad (7)$$

with $\langle V \rangle = b^3$. This value also represents exactly the fraction of hydrogen-bonded states $\langle h \rangle$ in equilibrium. As is is shown in Figure 2, with increasing box-size the transition to the plateau is shifting towards longer times.

However, the "Starr" correction according to Equation 1 accounts for the effect of periodic boundary conditions such that the "Luzar Chandler" and "Starr" approximations both envelope the true long-time limiting behaviour expressed by Equations 6 and 5. From Figure 2 it is obvious that the "Starr" approximation is representing the true (i.e. without PBCs) $C(t)$ worse than the "Luzar Chandler" approximation, since it is introducing a hardly qualifiable time-scaling behaviour for long times t, whereas the transition to a plateau in the "Luzar Chandler" approximation can be certainly better detected. Moreover, before transitioning to a plateau value, the "Luzar Chandler" approximation follows the true $C(t)$ more faithfully. To estimate the time-interval $t_{\max}$ up to which the computed $C(t)$ in presence of PBCs can be trusted, we determine the intersection of the $\lim_{t\to\infty} C(t)$ according to Equation 5 and the plateau-value C_∞ given by Equation 7:

$$t_{\max} = \frac{\langle V \rangle^{2/3}}{4\pi D' \cdot n^{2/3}}. \quad (8)$$

Here the contributions from the volume defining the hydrogen-bonded state cancel out each other, such that the time-interval $t_{\max}$ solely depends on the volume of the central box and the inter-diffusion coefficient of HB donors and acceptors D'. The true intersection time $t_{\max}$ is, of course, found for a value of $n = 1$ in Equation 8. However, as it is demonstrated in Figure 3, the deviation from the unperturbed HBPCF start to become significant at an earlier time, which can be taken into account

by a value of $n \approx 2$ for the "Luzar Chandler" approximation and about $n \approx 20$ for the "Starr" approximation. So, $t_{\max}$ is the time, when the HBPCF passes a threshold-value of $n \cdot C_\infty$. The corresponding values for $C(t_{\max})$ are indicated in Figure 3a for varying diffusion coefficients D' with a fixed box-size of $b = 6$, and in Figure 3b for box-sizes varying between $b = 4$ and $b = 20$ for a fixed diffusion coefficient of $D' = 1/6$.

B. A Correction-Term for Periodic Boundary Conditions

Now that we have established the time range up to which the computed HBPCFs can be trusted, we propose an *a posteriori* treatment of the long-time behaviour, correcting for the effect of periodic boundary conditions using "wrapped" trajectories. Obviously, it would be preferable, to have a correction-term, that provides the *true* (i.e. no PBCs) HBPCF. Our solution is based on the insight that such a correction needs to take into account the time structure of the correcting term. Here, we propose to describe the time-dependence in the following way:

$$C(t) = \frac{\langle h(0)h(t)\rangle}{\langle h\rangle} - \langle h\rangle\, s(t)\ , \qquad (9)$$

where $s(t)$ represents a "switching function", appropriately converting from the "Luzar-Chandler" approximation for short times $\lim_{t\to 0} s(t) = 0$ to the "Starr" correction at long times $\lim_{t\to\infty} s(t) = 1$ in such a way, that the true $t^{-3/2}$ long-time limiting behaviour is restored. To determine this switching function, we have computed $s(t)$ based on the data shown in Figure 3 according to

$$s(t) = C_\infty^{-1}\left[\frac{\langle h(0)h(t)\rangle}{\langle h\rangle} - C_{\text{no-pbc}}(t)\right]\ . \qquad (10)$$

Here, the expressions for $C_{\text{no-pbc}}(t)$ and C_∞ are given by Equations 6 and 7, respectively. The computed functions $s(t)$, of course, depend on the diffusion coefficients D' and the box-sizes b of the systems under consideration. For an universal, system-size- and particle-mobility-independent representation, we switch for $s(t)$ to a diffusion- and system-size-invariant time-scale. In Figure 4 the functions $s(t)$ for varying parameters D' and b are plotted on a reduced timescale u with $u = D't/b^2$. By doing so, all the computed functions $s(u)$ for different diffusion coefficients D' and box-sizes b fall on top of each other and are represented by a single master curve.

For deriving an analytical expression for the universal switching function $s(u)$, we analyse the behaviour of a hypothetical random walker on a periodic primitive cubic lattice with lattice spacing $d = (D'\delta t)^{1/2}$ for a given fixed time-step δt. Here the box size of the periodic unit cell is defined as k-times the lattice spacing $b = k \cdot d$. A peculiarity of a random walker on a lattice with a fixed time-step is that it can return to its origin only at an

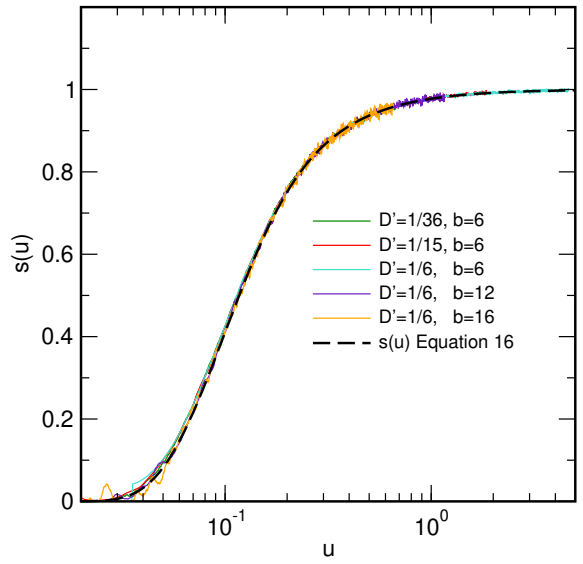

Figure 4. Universal representation of the switching functions $s(t)$ according to Equation 10, obtained for various diffusion coefficients D' and box-sizes b, plotted as a function of a reduced time-scale u with $u = D't/b^2$. The heavy dashed line indicates the exact analytical expression for the master curve $s(u)$ according to Equation 16.

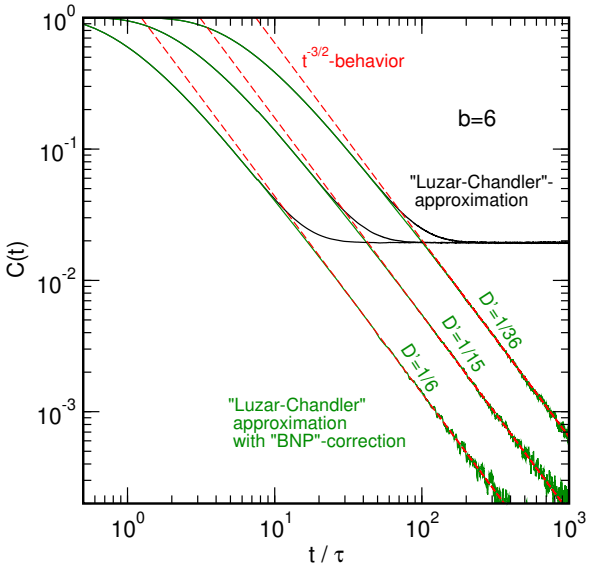

Figure 5. Hydrogen bond population correlation functions (HBPCFs) $C(t)$ for the "generic" random walker model with $D_{\text{jump}} = 1/6, 1/15$, and $1/36$. Solid black lines: $C(t)$ for a random walker with periodic boundary conditions with the indicated cubic box-lengths $b = 6$ computed according to Equation 3. Dashed red lines: Analytical solution for $\lim_{t\to\infty} C(t)$ according to Equation 5. Green solid lines: $C(t)$ computed from the data shown as black lines but applying the "BNP"-correction according to equations 16.

even number of time-steps, such that the number of lattice points that is accessible to the walker at either, even or uneven time steps is exactly one half of the total number of lattice points. Hence, the long-time limit of the corresponding HBPCF for a nonperiodic ($k \to \infty$) lattice follows

$$C(t) \approx \frac{1}{1/2\,\rho\,(4\pi D't)^{3/2}}, \tag{11}$$

where $\rho = d^{-3}$ is the density of lattice points. Typically, on-lattice random walkers follow this approximate behaviour already after a few time-steps. For the random walker on a periodic lattice, the fraction of the time the walker visits its starting point at $t = 0$, or its periodic images is hence given as

$$\langle h \rangle = C_\infty = \frac{2d^3}{b^3} = \frac{2}{k^3}\,. \tag{12}$$

Let us assume, the probability of finding a random walker at a certain point $p(\vec{r},t)$ follows a Gaussian distribution

$$p(\vec{r},t) = q(t)^{-1} e^{-|\vec{r}|^2/(4D't)} \tag{13}$$

with $q(t) = (4\pi D't)^{3/2}$. To compute the HBPCF within the "Luzar-Chandler" approximation, we have to consider contributions from all periodic images in addition to the central box. Hence,

$$C(t) = \frac{\langle h(0)h(t)\rangle}{\langle h\rangle} = \frac{2d^3}{q(t)} \times \sum_{n=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} e^{-b^2(n^2+m^2+l^2)/(4D't)}, \tag{14}$$

where the vectors (n,m,l) point at all periodic images, including the central unit cell located at (0,0,0). Following the definition for $s(t)$ from Equation 10, we have to subtract the contribution from the term for $(0,0,0)$ and multiply by $b^3/(2d^3)$, such that

$$s(t) = \frac{b^3}{q(t)}\left\{1 + 2\sum_{n=1}^{\infty}\exp\left[-\frac{b^2n^2}{4D't}\right]\right\}^3 - \frac{b^3}{q(t)}\,. \tag{15}$$

By substituting $u = D't/b^2$, we obtain

$$s(u) = \frac{1}{q(u)}\left\{1 + 2\sum_{n=1}^{\infty}\exp\left[-\frac{n^2}{4u}\right]\right\}^3 - \frac{1}{q(u)} \tag{16}$$

with $q(u) = (4\pi u)^{3/2}$. Here, Equation 16 shows a qualitatively correct limiting behaviour. For $u \to \infty$ the sum can be approximated as an integral over u, which converges to $(\pi u)^{1/2}$, leading to $\lim_{u\to\infty} s(u) = 1$. For $u \to 0$ the contribution from the individual exponentials are diminishing, such that $\lim_{u\to 0} s(u) = 0$. Unfortunately, the sum in Equation 16 has to be computed numerically. However, for the relevant range of u used in molecular simulations with $u < 10^2$, the sum converges quickly, such that a $n_{\max} \approx 100$ is more than sufficient. For very small systems (e.g. defined by a small value of $k \leq 4$) the Gaussian distribution used in Equation 13 should be replaced by a multinomial distribution. For system sizes of $k \geq 10$, however, the results from Equation 13 and lattice simulations are practically indistinguishable and for $4 < k < 10$ it serves as a very good approximation. Therefore, as an application to molecular simulations, the use of a Gaussian distribution is perfectly adequate.

We would like to point out that the application of Equation 16 to non-cubic periodic boundary conditions is straightforward: We just need to replace the cubed sum in Equation 16 by a product of three sums, where u is computed for each direction x, y, z separately using $u_\alpha = D'_\alpha t/b_\alpha^2$ with $\alpha \in \{x,y,z\}$. The normalization factor has to be adjusted accordingly with $q(u_x,u_y,u_z) = 8\pi^{3/2}(u_x u_y u_z)^{1/2}$. This procedure, of course, also incorporates non-isotropic systems, where variations of the diffusion coefficient in different directions are observed, for example for a smectic phase of a liquid crystal, or for the lateral diffusion within a membrane.

Finally, in Figure 4 we compare the prediction of $s(u)$ according to Equation 16 with the corresponding functions derived from the random walker simulations. In Figure 5, we apply the correction outlined in Equations 16, using the HBPCF-data from a random walker model with $b = 6$ and varying diffusion coefficients D' based on the "Luzar-Chandler" approximation as input source. As indicated by the solid green lines depicted in Figure 5, the proposed time-dependent "BNP" (Busch, Neumann, Paschek) correction is able to fully recover the correct long-time limiting behaviour of the computed HBPCFs, thus being superior to any of the other hitherto employed treatments.

C. HBPCFs for Liquid Water as a Test Case

A rigorous definition of a hydrogen-bonded state would need to involve electronic structure calculations [46–48]. Since quantum mechanical calculations are not yet feasible at the required scale, the most rational approach is to rely on energetical or geometrical criteria [16, 49]. Here, we follow the procedure of Kumar et al. [49] to identify a hydrogen-bonded state as a basin on a "free energy landscape" of suitable geometric parameters. As criteria for a hydrogen bond in liquid TIP4P/2005 water we use an intermolecular O···H distance $r_{\mathrm{OH}} \leq 0.25\,\mathrm{nm}$ and a cosine of the angle α between the intermolecular O···H vector and the intramolecular O–H bond-vector donating the HB of $\cos(\alpha) \leq -0.6$, both encompassing the corresponding "free energy basin". We would like to emphasise that Kikutsuji et al. [50] have recently explored numerous different geometrical HB definitions for TIP4P/2005 water based on Kumar's suggestions, and found that they were all leading to consistent results.

We have examined MD simulations of TIP4P/2005 water at 273 K and 298 K of 10 ns length. Details of the

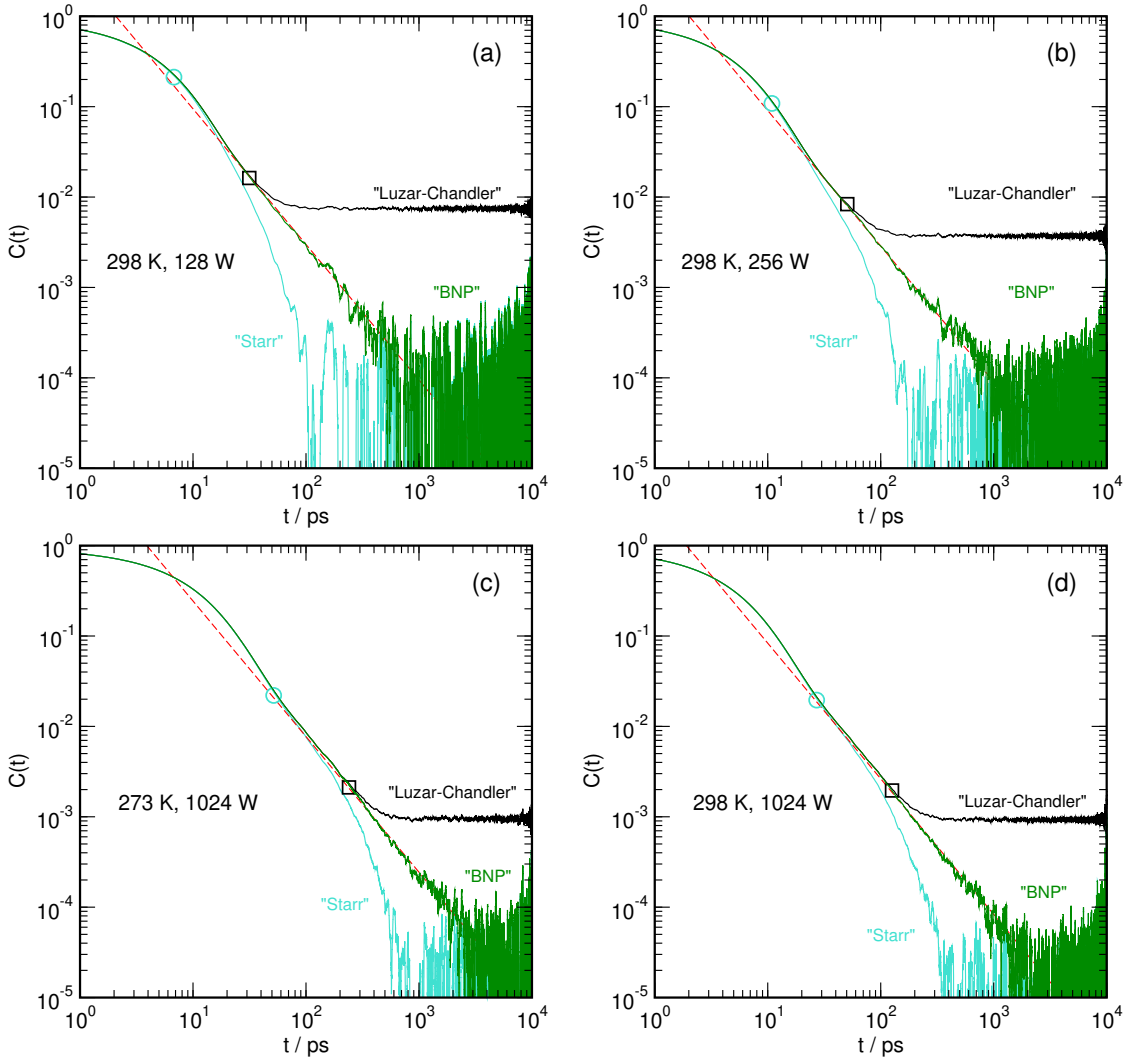

Figure 6. Hydrogen bond population correlation functions for TIP4P/2005 water at 273 K and 298 K for system sizes with $N=128$, $N=256$, and $N=1024$ water molecules: a) 128 water molecules at 298 K. b) 256 water molecules at 298 K. c) 1024 water molecules at 273 K. d) 1024 water molecules at 298 K. Dashed red lines: analytical representation for $\lim_{t\to\infty} C(t)$ according to Equation 5 using the parameters given in Table **??**. The open symbols indicate the times $t_{\max}$ according to Equation 8 for the "Luzar-Chandler" approximation with $n=2$ and the "Starr" correction with $n=20$. Solid black lines: HBPCFs according to the "Luzar-Chandler" approximation (Equation 3). Solid turqoise lines: HBPCFs according to the "Starr" correction (Equation 1). Solid green lines: HBPCFs according to the "BNP" correction (Equation 16) using the parameters given in Table **??**. Note that water serves as both, donor and acceptor, hence $D' = 2 \times D_{\text{self}}$.

analysis of our MD simulations are summarised in Table **??**. Note the system-size dependence of the self-diffusion coefficient, as shown in Table **??**. This system size dependence behaves exactly like it has been reported earlier from simulations of TIP3P water [51], and its origin has been linked to hydrodynamic self-interactions [51, 52], and can be explained quantitatively. Based on the above-defined HB criteria, Figure **??** shows the computed hydrogen bond population correlation functions for TIP4P/2005 water. The obtained plateau-values for the "Luzar-Chandler" approximations are given also in Table **??**. The plateau values C_∞ correspond roughly to the inverse number of water molecules in the system, but not quite exactly, since not all the water molecules are engaged in a hydrogen bond all the time. Here, an increasing fraction of broken HBs adds to the time-slice hydrogen bonded partners cannot spend with each other, and is therefore leading to a lower C_∞. This way we can explain the small but consistent temperature dependence of C_∞ observed in Table **??**. Similar to the behaviour of the "generic" random walker model, both the "Luzar-Chandler" and "Starr" approximations are enveloping the $t^{-3/2}$ long-time limiting behavior indicated by red dashed lines based on Equation 5 using the parameters given in Table **??**. By applying the "BNP" correction to the "Luzar–Chandler" data using the box size-, self-

Table I. Parameters describing the MD simulations performed under NVT conditions at the indicated densities ρ. b: MD unit cell box-length. ρ_{acc}: hydrogen bond acceptor density. s: scaling parameter used in Equation 5. D_{self}: water self diffusion coefficient, determined from the slope of the center-of-mass mean square displacement of the water molecules. C_∞: averaged plateau value of the HBPCFs obtained for the "Luzar-Chandler" approximation. τ_{HB}: HB lifetime obtained by numerically integrating the respective HBPCFs. St: "Starr". BNP: "BNP".

N	T/K	ρ/g cm^{-3}	b/nm	ρ_{acc}/nm^{-3}	s	$D_{\mathrm{self}}/10^{-9}$ m^{2}s^{-1}	C_∞	$\tau_{\mathrm{HB,St}}$/ps	$\tau_{\mathrm{HB,BNP}}$/ps
128	273	0.9997	1.56462	33.42	1.00	0.92	7.6×10^{-3}	9.52	11.98
256	273	0.9997	1.97130	33.42	0.98	0.97	3.8×10^{-3}	10.01	11.88
1024	273	0.9997	3.12924	33.42	0.96	1.01	9.5×10^{-4}	11.00	12.01
128	298	0.9972	1.56597	33.33	1.00	1.85	7.5×10^{-3}	4.75	5.81
256	298	0.9972	1.97300	33.33	0.98	1.93	3.7×10^{-3}	4.82	5.78
1024	298	0.9972	3.13194	33.33	0.96	2.07	9.3×10^{-4}	5.30	5.76

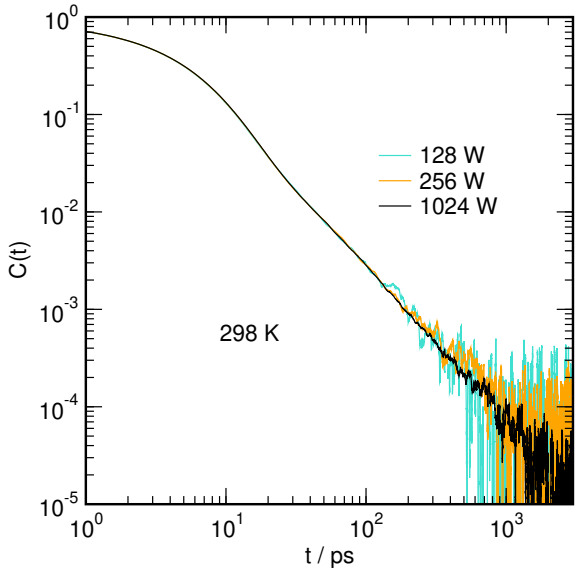

Figure 7. Hydrogen bond population correlation functions for TIP4P/2005 water at 298 K for system sizes with $N = 128$, $N = 256$, and $N = 1024$ water molecules using the "BNP" correction.

diffusion-, and C_∞-data given in Table **??**, we are able to fully restore the long-time limiting behaviour in a correct fashion, independently of the chosen temperature or system size. Here, we are showing the computed HBPCFs over the entire simulation time-range, reporting the value of the $C(t)$ function over five orders of magnitude. However, depending on the system size, the restored HBPCFs are dominated by noise for times greater than about one nanosecond. Here, the 1024-molecule systems exhibit visibly better statistics. For a better comparison, the computed HBPCFs including the "BNP" correction for water at 298 K and different system sizes are superimposed in Figure 7. Note that all graphs are perfectly aligned and lie on top of each other. The only noticeable difference is the decreasing noise at large t with increasing system size. At this point we can only speculate that a mechanism similar to the one that keeps the computed shear viscosity from being system-size-dependent [51] is at work here as well. Hydrogen bond lifetimes, computed as integrals over HBPCFs represent a simple and popular means to quantify HB kinetics. To study the effect of the "BNP" correction on the HB lifetimes, we have used "brute-force" numerical integration of the unaltered HBPCFs according to the Starr approximation over the whole available time-interval. Since we know the long-time limiting behavior of the HBPCFs according to the "BNP" correction, we numerically integrate the HBPCS up to a time $t^* = 200$ ps and correct for the long-time limit according to

$$\tau_{\mathrm{HB,BNP}} = \int_0^{t^*} C(t)\,dt + \frac{1}{s\,\rho_{\mathrm{acc}} 4(\pi D')^{3/2}\sqrt{t^*}}\,. \qquad (17)$$

As shown in Table **??**, the lifetimes according to the "Starr" approximation show a clear system size dependence. This behaviour can be well understood, since with increasing system-size the deviation from the unperturbed behaviour is starting at a later time. The lifetimes according to the "BNP" correction are consistent and show no clear system-size dependence in accordance with the observed system-size independence of the computed HBPCFs shown in Figure 7. Their values obviously seem to represent the limiting values the "Starr" approximation data aim towards for infinite system sizes.

IV. CONCLUSIONS

In this contribution we have analysed the effect of periodic boundary conditions (PBCs) on the long-time limiting behaviour of hydrogen bond population correlation functions (HBPCFs) obtained from untreated "wrapped" trajectories. Here the breaking and re-formation of hydrogen bonds serves as a test case scenario for the more general application of the reversible diffusion-influenced geminate recombination, which could be applied in the form of similar population correlation functions (PCFs) to related problems such as diffusion controlled ion-pair

re-formation [53]. In order to compare HPBCFs in presence of PBCs with the true unperturbed behaviour, we employed a simple random walker model, allowing to easily vary the size of the periodic box and the diffusion coefficient. Here, the hydrogen bonded state is defined by a sphere of a given radius around the starting point of the walker at $t = 0$. The HBPCFs for the corresponding random walker in absence of PBCs are given as a simple analytical expression. For defining HBPFs, essentially two different strategies have been used, either following Starr et al. [16] or Luzar and Chandler [8]. The "Starr" approximation has the advantage of properly decaying to zero, whereas the "Luzar-Chandler" definition follows more closely the behaviour in absence of PBCs, but is finally transitioning to a plateau value. Here we demonstrate that each method is deviating in a different fashion from the true long-time limiting behaviour, but enveloping the unperturbed function. By quantitatively analysing the random walker trajectories, however, we find that the "Luzar-Chandler"-approximation is following the true (i.e. no PBCs) behaviour up to one order of magnitude in time more faithfully. Based on an expression for the long-time limiting behaviour of the HBPCFs following a $t^{-3/2}$ time-dependence, a simple expression is given for estimating the maximum time interval up to which the computed HBPCFs for either of the two approximations are reliably describing the long-time limiting behaviour and can therefore be trusted. An exact *a posteriori* correction for systems with periodic boundary conditions is derived, which can be easily computed and which is able to fully recover the true $t^{-3/2}$ long-time behaviour for HBPCFs based on data computed using the "Luzar-Chandler"-approximation. The suggested "BNP" correction just requires the knowledge of the HB donor-acceptor inter-diffusion coefficient and the size of the box-dimensions. For comparison, we have computed HBPCFs for liquid water at 273 K and 298 K from MD simulations of TIP4P/2005 model water for varying system sizes and temperatures using this newly introduced correction. Quite interestingly, the computed HBPCFs do not show indications of a system-size dependence, despite the fact that the self-diffusion coefficients is system-size dependent.

Finally, although it is true that the same HBPCFs could be computed using the trajectory "unwrapping" procedure described by Markovitch and Agmon[18], we believe that the BNP-approach is of value, since it just requires a simple post-processing procedure which can be realized in the form of a Perl-script with a few lines of code. Even then the post-processing typically takes less than a second for time correlation data sets with multiple ten-thousand entries. Another argument that could be made in favour of the BNP approach, we think, could be numerical efficiency, since the maximum number of correlation functions that need to be computed is limited by the combination of all possible donor-acceptor pairs in the system. On the other hand with respect to the "unwrapping" approach: for a truly periodic system, the number of possible donor-acceptor combinations is infinite. Even the number of combinations that meet (i.e. form a HB) at least once is infinite. To achieve the same statistical accuracy as the BNP approach, one would therefore have to identify all unique (non-duplicate) donor-acceptor combinations of all periodic images that meet at least once, which we suspect, could be larger than the number of donor-acceptor pairs in the system.

ACKNOWLEDGEMENTS

We thank the computer center at the University of Rostock (ITMZ) for providing and taking care of computational resources. We also thank the Leibniz Association, the State of Mecklenburg-Vorpommern, and the University of Rostock for financial support within the ComBioCat programme.

DATA AVAILABILITY STATEMENT

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

[1] R. Ludwig. Water: From clusters to the bulk. *Angew. Chem. Int. Ed.*, 40:1808–1827, 2001.
[2] F. Stillinger. Water revisited. *Science*, 209:451–457, 1980.
[3] B. M. Ladanyi and M. S. Skaf. Computer-simulation of hydrogen-bonding liquids. *Annu. Rev. Phys. Chem*, 44:335–368, 1993.
[4] P. A. Hunt, C. R. Ashworth, and R. P. Matthews. Hydrogen bonding in ionic liquids. *Chem. Soc. Rev.*, 44:1257–1288, 2015.
[5] G. A. Jeffrey and W. Saenger. *Hydrogen Bonding in Biological Structures*. Springer, 1991.
[6] J. D. Watson and F. H. C. Crick. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.
[7] A. Luzar and D. Chandler. Effect of environment on hydrogen bond dynamics in liquid water. *Phys. Rev. Lett.*, 76(6):928–931, 1996.
[8] A. Luzar and D. Chandler. Hydrogen-bond kinetics in liquid water. *Nature*, 379(6560):55–57, 1996.
[9] A. Luzar. Resolving the hydrogen bond dynamics conundrum. *J. Chem. Phys.*, 113:10663–10675, 2000.
[10] F. H. Stillinger. Theory and molecular models for water. *Adv. Chem. Phys.*, 31:1–101, 1975.
[11] D. C. Rapaport. Hydrogen bonds in water: Network organization and lifetimes. *Mol. Phys.*, 50(5):1151–1162, 1983.

[12] A. Geiger, P. Mausbach, J. Schnitker, R. L. Blumberg, and H. E. Stanley. Structure and dynamics of the hydrogen-bond network in water by computer-simulations. *J. Phys. (France)*, 45:C13–C30, 1984.

[13] D. A. Zichi and P. J. Rossky. Solvent molecular-dynamics in regions of hydrophobic hydration. *J. Chem. Phys.*, 84:2814–2822, 1986.

[14] F. Sciortino and S. L. Fornilil. Hydrogen-bond cooperativity in simulated water - time-dependence analysis of pair interactions. *J. Chem. Phys.*, 90:2786–2792, 1989.

[15] F. Sciortino, P. H. Poole, H. E. Stanley, and S. Havlin. Lifetime of the bond network and gel-like anomalies in supercooled water. *Phys. Rev. Lett.*, 61:1686–1689, 1990.

[16] F. W. Starr, J. K. Nielsen, and H. E. Stanley. Hydrogen-bond dynamics for the extended simple point-charge model of water. *Phys. Rev. E*, 62:579–587, 2000.

[17] D. van der Spoel, P. J. van Maaren, P. Larsson, and N. Tîmneanu. Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media. *J. Phys. Chem. B*, 110:4393–4398, 2006.

[18] O. Markovitch and N. Agmon. Reversible geminate recombination of hydrogen-bonded water molecule pair. *J. Chem. Phys.*, 129:084505, 2008.

[19] N. Agmon and G. H. Weiss. Theory of non-markovian reversible dissociation reactions. *J. Chem. Phys.*, 91:6937–6942, 1989.

[20] D. Laage and J. T. Hynes. A molecular jump mechanism of water reorientation. *Science*, 311:832–835, 2006.

[21] D. Laage and J. T. Hynes. On the molecular mechanism of water reorientation. *J. Phys. Chem. B*, 112:14230–14242, 2008.

[22] D. C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, Cambridge, UK, 2nd edition, 2004.

[23] A. Strate, J. Neumann, V. Overbeck, A.-M. Bonsa, D. Michalik, D. Paschek, and R. Ludwig. Rotational and translational dynamics and their relation to hydrogen bond lifetimes in an ionic liquid by means of NMR relaxation time experiments and molecular dynamics simulation. *J. Chem. Phys.*, 148(19):193843, 2018.

[24] J. Neumann, D. Paschek, A. Strate, and R. Ludwig. Kinetics of hydrogen bonding between ions with opposite and like charges in hydroxyl-functionalized ionic liquids. *J. Phys. Chem. B*, 125:281–286, 2021.

[25] J. L. F. Abascal and C. Vega. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.*, 123:234505, 2005.

[26] C. Vega and J. L. F. Abascal. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.*, 13:19633–19688, 2011.

[27] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. GROMACS: fast, flexible, and free. *J. Comput. Chem.*, 26(16):1701–1718, 2005.

[28] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4(3):435–447, 2008.

[29] S.Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.*, 52:255–268, 1984.

[30] W. G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695–1697, 1985.

[31] U. Essmann, L. Petera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8593, 1995.

[32] C. L. Wennberg, T. Murtola, B. Hess, and E. Lindahl. Lennard-jones lattice summation in bilayer simulations has critical effects on surface tension and lipid properties. *J. Chem. Theory Comput.*, 9:3527–3537, 2013.

[33] C. L. Wennberg, T. Murtola, S. Páll, M. J. Abraham, B. Hess, and E. Lindahl. Direct-space corrections enable fast and accurate lorentz-berthelot combination rule lennard-jones lattice summation. *J. Chem. Theory Comput.*, 11:5737–5746, 2015.

[34] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *J. Comput. Chem.*, 13:952–962, 1992.

[35] T. S. Grigera. Everything you wish to know about correlations but are afraid to ask. `arXiv:2002.01750v1`, `[cond-mat.stat-mech]`, 2020.

[36] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, USA, 2 edition, 1992.

[37] N. Michaud-Agraval, E. J. Denning, T. B. Woolf, and O. Beckstein. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.*, 32:2319–2327, 2011.

[38] R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, Jan Domański, D. L. Dotson, S. Buchoux, I. M. Kenney, and O. Beckstein. MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations. In S. Benthall and S. Rostrup, editors, *Proceedings of the 15th Python in Science Conference*, pages 98–105, Austin, TX, 2016.

[39] T. Oliphant. NumPy: A guide to NumPy. USA: Trelgol Publishing, 2006–.

[40] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[41] N. Agmon, E. Pines, and D. Huppert. Geminate recombination in proton-transfer reactions .2. comparison of diffusional and kinetic schemes. *J. Chem. Phys.*, 88:5631–5638, 1988.

[42] N. Agmon and A. Szabo. Theory of reversible diffusion-influenced reactions. *J. Chem. Phys.*, 92:5270–5284, 1990.

[43] H. Chen, G. A. Voth, and N. Agmon. Kinetics of proton migration in liquid water. *J. Phys. Chem. B*, 114:333–339, 2010.

[44] M. E. Tuckerman, A. Chandra, and D. Marx. A statistical mechanical theory of proton transport kinetics in hydrogen-bonded networks based on population correlation functions with applications to acids and bases. *J. Chem. Phys.*, 133:124108, 2010.

[45] P. W. Atkins and J. de Paula. *Physical Chemistry*. F. H. Freeman and Company, New York, 9th edition, 2010.

[46] F. Weinhold and R. A. Klein. What is a hydrogen bond? resonance covalency in the supramolecular domain. *Chem. Educ. Res. Pract.*, 15:276–285, 2014.

[47] S. J. Grabowski. What is the covalency of hydrogen bonding? *Chem. Rev.*, 111:2597–2625, 2011.

[48] F. Weinhold and C. R. Landis. *Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective*. Cambridge University Press, 2005.

[49] R. Kumar, J. R. Schmidt, and J. L. Skinner. Hydrogen bonding definitions and dynamics in liquid water. *J. Chem. Phys.*, 126(20):204107, 2007.

[50] T. Kikutsuji, K. Kim, and N. Matubayasi. Consistency of geometrical definitions of hydrogen bonds based on the two-dimensional potential of mean force with respect to the time correlation in liquid water over a wide range of temperatures. *J. Mol. Liq.*, 294:111603, 2019.

[51] I.-C. Yeh and G. Hummer. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B*, 108:15873–15879, 2004.

[52] B. Dünweg and K. Kremer. Molecular dynamics simulation of a polymer chain in solution. *J. Chem. Phys.*, 99:6983–6997, 1993.

[53] S. Gehrke and B. Kirchner. Robustness of the hydrogen bond and ion pair dynamics in ionic liquids to different parameters from the reactive flux method. *J. Chem. Eng. Data*, 65(3):1146–1158, 2020.