

# Unwrapping NPT Simulations to Calculate Diffusion Coefficients

Jakob Tómas Bullerjahn,\* Sören von Bülow, Maziar Heidari, Jérôme Hénin, and Gerhard Hummer\*



Cite This: *J. Chem. Theory Comput.* 2023, 19, 3406–3417



Read Online

ACCESS |



Metrics & More

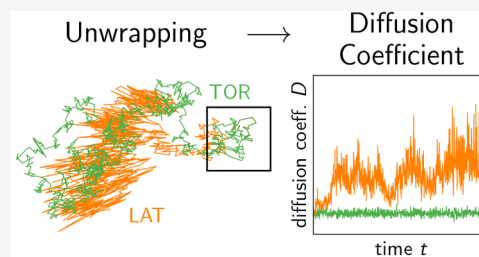


Article Recommendations



Supporting Information

**ABSTRACT:** In molecular dynamics simulations in the NPT ensemble at constant pressure, the size and shape of the periodic simulation box fluctuate with time. For particle images far from the origin, the rescaling of the box by the barostat results in unbounded position displacements. Special care is thus required when a particle trajectory is unwrapped from a projection into the central box under periodic boundary conditions to a trajectory in full three-dimensional space, e.g., for the calculation of translational diffusion coefficients. Here, we review and compare different schemes in use for trajectory unwrapping. We also specify the corresponding rewinding schemes to put an unwrapped trajectory back into the central box. On this basis, we then identify a scheme for the calculation of diffusion coefficients from NPT simulations, which is a primary application of trajectory unwrapping. In this scheme, the wrapped and unwrapped trajectory are mutually consistent and their statistical properties are preserved. We conclude with advice on best practice for the consistent unwrapping of constant-pressure simulation trajectories and the calculation of accurate translational diffusion coefficients.



## 1. INTRODUCTION

Molecular dynamics (MD) simulations are performed by numerically solving the classical equations of motion for every particle in a given system. For systems in the condensed phase, such as proteins in water, these simulations are usually conducted in volumes of finite size subject to periodic boundary conditions (PBCs). In constant-volume simulations, one can think of the periodic system either as a single box in which opposite faces are identified under what are also referred to as toroidal boundary conditions, or as an infinite periodic lattice of replicates of the central simulation box. In the *toroidal view*, a particle leaving the central simulation box placed at the coordinate origin re-enters the box at the opposing face, as it would when moving around on a torus. In the *lattice view*, each particle corresponds to a collection of infinitely many points on a periodic lattice, whose lattice constants are determined by the box size and shape. The toroidal view naturally leads to so-called wrapped trajectories, where particles at every instance in time are contained within the central box (and positions outside the box do not make mathematical sense). By contrast, in the lattice view each individual marked point on the lattice representing a particular particle can traverse the full three-dimensional space, resulting in an associated unwrapped trajectory. For simulation boxes of constant volume in constant-energy (NVE) and constant-temperature (NVT) ensembles, the task of unwrapping a trajectory therefore corresponds to transforming from the toroidal view to the lattice view.

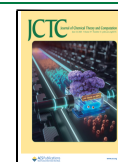
In constant-pressure (NPT) simulations, however, the task of unwrapping becomes somewhat ambiguous, because the barostat constantly modifies the size and shape of the simulation box to keep the average pressure fixed. The positions of the

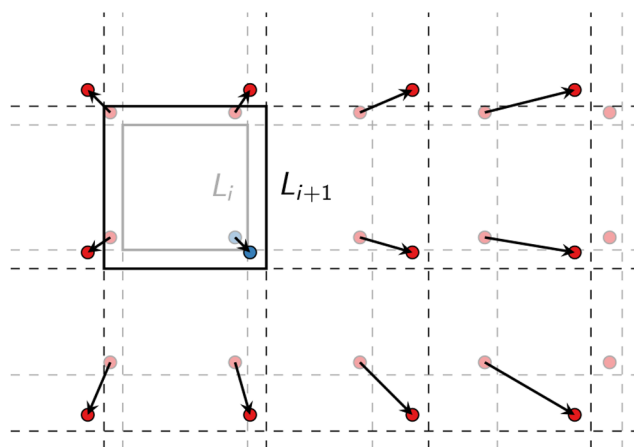
particles within the box thereby get rescaled.<sup>1</sup> In the lattice view of PBCs, the periodic lattice is now fluctuating. Importantly, the motion of particles purely as a result of the barostat action depends on their distance from the central simulation box and is thus unbounded (see Figure 1). By contrast, in the toroidal view particles stay in the box with effectively bounded displacements caused by barostat position rescaling. These differences between the toroidal and lattice views seem to have caused some confusion, as there are at least three different algorithms currently in use to unwrap trajectories of constant-pressure MD simulations.

Here, we review and compare the different schemes proposed for trajectory unwrapping at constant pressure (section 2). We use analytic calculations and numerical examples to demonstrate that lattice-preserving unwrapping schemes give rise to unwrapped trajectories with exaggerated fluctuations when used to unwrap NPT simulation data. In extreme cases, the dynamics of these unwrapped trajectories differs sharply from the dynamics of the associated wrapped trajectories (sections 3 and 5). As a consequence, diffusion coefficient estimates are compromised, an effect that becomes apparent already for bulk water at ambient conditions simulated in the NPT ensemble over a microsecond time scale. By contrast, we find that a recently proposed off-lattice unwrapping scheme<sup>2</sup> preserves the

Received: March 17, 2023

Published: May 31, 2023





**Figure 1.** Barostat box rescaling in lattice view of PBCs. In the lattice view, the displacement resulting from barostat-induced rescaling of the box volume grows with the distance from the reference box centered at the coordinate origin. The central boxes before and after barostat action are indicated by gray and black squares, respectively, and the corresponding periodic images of a particle by circles with faint and solid colors. As a result of barostat rescaling alone, particle images (red) away from the central box move farther than the reference particle (blue), as indicated by the arrows.

statistical properties of the wrapped trajectory and should therefore be preferred for the calculation of translational diffusion coefficients. However, because the scheme does not adhere to the lattice view, it does not preserve distances.<sup>3</sup> Molecules should thus first be made “whole” and then unwrapped, e.g., according to their center of mass. We conclude by giving guidance to practitioners on how to extract reliable diffusion coefficient estimates from constant-pressure MD simulations (sections 5.5 and 6).

## 2. UNWRAPPING ALGORITHMS

**2.1. Heuristic Lattice-View (HLAT) Scheme.** Some MD simulation and visualization software packages implement a lattice-preserving unwrapping scheme (see, e.g., `trjconv` in GROMACS<sup>4</sup> and `cpptraj` in Ambertools<sup>5</sup>), which in one dimension (1D) can be cast into the following form:

$$u_{i+1}^{\text{HLAT}} = w_{i+1} - \left\lfloor \frac{w_{i+1} - u_i^{\text{HLAT}}}{L_{i+1}} + \frac{1}{2} \right\rfloor L_{i+1}. \quad (1)$$

Here,  $w_i$  denotes the wrapped position of a particle inside the simulation box of width  $L_i$  at integration step  $i$  corresponding to time  $t_i$ ,  $u_i^{\text{HLAT}}$  is the corresponding unwrapped position predicted by the HLAT scheme (called the “heuristic scheme” in refs 2 and 3), and  $\lfloor \cdot \rfloor$  denotes the floor function. This scheme defines the unwrapped position at time  $i + 1$  as the particular lattice image of the wrapped position that minimizes the unwrapped displacement from time  $i$  to  $i + 1$ , making it intuitively appealing. In ref 2, however, it was shown that in simulations at constant pressure the above scheme occasionally unwraps particles into the wrong box, which results in an artificial speed up of the particles. This observation was later confirmed in ref 3.

**2.2. Toroidal-View-Preserving (TOR) Scheme.** After exposing the shortcomings of the HLAT scheme, three of the authors of the present paper proposed an alternative unwrapping

scheme, which resolves the issues of eq 1 and translates to the following evolution equation in 1D:<sup>2</sup>

$$u_{i+1}^{\text{TOR}} = u_i^{\text{TOR}} + (w_{i+1} - w_i) - \left\lfloor \frac{w_{i+1} - w_i}{L_{i+1}} + \frac{1}{2} \right\rfloor L_{i+1}. \quad (2)$$

Taking a toroidal view of PBCs, the TOR scheme considers minimal displacement vectors within the simulation box, which are added together to form an unwrapped trajectory. By design, it therefore preserves the dynamics of the wrapped trajectory. However, the TOR scheme should only be used to unwrap the trajectories of single particles, such as the center of mass of a molecule or a well-chosen reference atom. If the scheme is applied separately to multiple atoms of the same molecule, whose intramolecular bonds cross the periodic boundaries, then the atoms in question get incorrectly displaced with respect to each other, resulting in an unphysical stretching of the bonds connecting them together.<sup>3</sup> Therefore, molecules should first be made “whole” and then unwrapped.

We note that the TOR scheme (eq 2) appeals to the theory of diffusion coefficients  $D$  in terms of the autocorrelation function of the velocity  $v(t)$ , which in one dimension leads to the Green–Kubo relation<sup>6</sup>  $D = \lim_{\epsilon \rightarrow 0^+} \int_0^\infty dt \exp(-\epsilon t) \langle v(t)v(0) \rangle$ . From this expression, one obtains the Einstein relation  $D = \lim_{t \rightarrow \infty} \langle [u(t) - u(0)]^2 \rangle / (2t)$  by writing the position  $u(t)$  as a sum of infinitesimal displacements,  $u(t) = u(0) + \int_0^t dt' v(t')$ . In analogy, the TOR scheme (eq 2) adds up the minimal displacements between saved configurations to construct an unwrapped trajectory  $u(t)$ .

**2.3. Modern Lattice-View (LAT) Scheme.** An alternative to the HLAT scheme, which takes a lattice view of PBCs without succumbing to the known shortcomings of HLAT, is implemented in the `qwrap`<sup>7</sup> software package. To our knowledge, this scheme was never explicitly documented in the literature prior to implementation, but the LAMMPS simulation software<sup>8</sup> uses it to write out unwrapped coordinates.

In the lattice view, crossing the periodic boundaries corresponds to shifting the identity of the particle in the central box to one of its lattice images. The LAT unwrapping scheme keeps track of these shifts using integer image numbers  $n_i$  that indicate how many periodic images the current wrapped coordinates are away from the original, unwrapped particle. The image number  $n_i$  can be obtained either by explicit bookkeeping of image changes due to wrapping (as done by the `remap` function of LAMMPS), or by detecting large jumps in the wrapped coordinates (as done by the `gunwrap` feature of `qwrap`). In both cases, the unwrapped coordinate can be obtained as a lattice image of its wrapped counterpart, i.e.,

$$u_i^{\text{LAT}} = w_i - n_i L_i, \\ n_{i+1} = \sum_{j=0}^i \left\lfloor \frac{w_{j+1} - w_j}{L_{j+1}} + \frac{1}{2} \right\rfloor. \quad (3)$$

This is done in `qwrap`, and in LAMMPS whenever unwrapped coordinates are necessary, such as for output or for use by the Colvars library.<sup>9</sup>

Recently, Kulke and Vermaas<sup>3</sup> proposed a correction to the TOR scheme with the aim to preserve the underlying lattice structure. Their scheme takes the following form in 1D:

$$u_{i+1}^{\text{LAT}} = u_i^{\text{LAT}} + (w_{i+1} - w_i) - \left[ \frac{w_{i+1} - w_i}{L_{i+1}} + \frac{1}{2} \right] L_{i+1} - \left[ \frac{w_i - u_i^{\text{LAT}}}{L_i} + \frac{1}{2} \right] (L_{i+1} - L_i). \quad (4)$$

However, in hindsight it turns out that eq 4 is equivalent to the earlier LAT scheme (eq 3). This can be seen by substituting eq 3 into eq 4, giving

$$\begin{aligned} u_{i+1}^{\text{LAT}} &= w_{i+1} - n_i L_i - \left[ \frac{w_{i+1} - w_i}{L_{i+1}} + \frac{1}{2} \right] L_{i+1} \\ &\quad - \left[ n_i + \frac{1}{2} \right] (L_{i+1} - L_i) \\ &= w_{i+1} - \left[ \frac{w_{i+1} - w_i}{L_{i+1}} + \frac{1}{2} \right] L_{i+1} - n_i L_{i+1} \\ &= w_{i+1} - n_{i+1} L_{i+1}, \end{aligned}$$

where we exploited the relation  $[x + n] = [x] + n$  with  $n \in \mathbb{Z}$ , an integer number in the second step. For this reason, we make no distinction between eqs 3 and 4 and refer to them both as the LAT scheme.

In what follows, we restrict our discussion to the comparison of the TOR and LAT schemes, as the HLAT scheme has already been established as faulty.

### 3. THEORY

Here, we describe a minimal stochastic model of a diffusive particle inside a fluctuating box with PBCs, which we use to generate numerical data and to highlight the differences between the unwrapping schemes via analytic calculations. We also develop and identify appropriate (re)wrapping schemes for the TOR and LAT schemes, respectively. Finally, we derive an upper bound for the frequency with which particle coordinates should be sampled to make sure that all boundary crossings are accounted for.

**3.1. Minimal Stochastic Model.** The 1D Gaussian model was introduced in ref 2 and provides a minimal theoretical description of constant-pressure MD simulations. It consists of a Wiener process  $w$  that evolves between two periodic boundaries, located at  $\pm L_i/2$  at time integration step  $i$ , which are themselves modeled as Gaussian white noise. Due to box length fluctuations, the value of the process gets rescaled in each time step, after which a diffusive displacement is performed. The model gives rise to the following wrapped trajectory:

$$\begin{aligned} w_{i+1} &= w_i + \left( \frac{L_{i+1}}{L_i} - 1 \right) w_i + \sigma_w R_{i+1} \\ &\quad - \left[ \frac{w_i}{L_i} + \frac{\sigma_w R_{i+1}}{L_{i+1}} + \frac{1}{2} \right] L_{i+1}, \\ L_{i+1} &= \bar{L} + \sigma_L S_{i+1}, \end{aligned} \quad (5)$$

where  $R_i, S_i \sim \mathcal{N}(0, 1)$  denote uncorrelated normally distributed random variables with zero mean and unit variance,  $\bar{L}$  is the average length of the 1D simulation box, and  $\sigma_w$  and  $\sigma_L$  determine the noise amplitudes of the random processes driving particle diffusion and box fluctuations, respectively.

In the absence of wrapping events, the displacements resulting from box rescaling and diffusion would give rise to a trajectory of the following form:

$$u_{i+1} = u_i + \left( \frac{L_{i+1}}{L_i} - 1 \right) w_i + \sigma_w R_{i+1}, \quad (6)$$

which can be regarded as the unwrapped partner trajectory to  $w$  of eq 5. Note that the second term on the right-hand side of eq 6 represents multiplicative noise, as can best be seen in the limit  $\sigma_L \ll \bar{L}$ , where we have  $L_{i+1}/L_i - 1 = \sqrt{2} \sigma_L S'_{k+1}/\bar{L} + O(\sigma_L^2/\bar{L}^2)$  with  $S'_{k+1} = (S_{k+1} - S_k)/\sqrt{2} \sim \mathcal{N}(0, 1)$ . Notably, this multiplicative-noise term is also present in eq 5, which makes both  $w$  and  $u$  distinctively different from an ordinary Wiener process. Yet, because the noise amplitude is only proportional to  $w_i$  (and not  $u_i$ ), it remains bounded and does not overshadow the diffusive process.

### 3.2. Differences Between Unwrapping Schemes.

Unwrapping the wrapped trajectory of eq 5 using the TOR scheme results in an unwrapped trajectory that coincides with eq 6. This can be demonstrated by substituting eq 5 with  $n = [w_i/L_i + \sigma_w R_{i+1}/L_{i+1} + 1/2] \in \mathbb{Z}$  into eq 2, giving

$$\begin{aligned} u_{i+1}^{\text{TOR}} &= u_i^{\text{TOR}} + \left( \frac{L_{i+1}}{L_i} - 1 \right) w_i + \sigma_w R_{i+1} \\ &\quad - \left[ \frac{w_i}{L_i} - \frac{w_i}{L_{i+1}} + \frac{\sigma_w R_{i+1}}{L_{i+1}} + \frac{1}{2} \right] L_{i+1}. \end{aligned}$$

The remaining floor function evaluates to zero as long as  $\sigma_w R_{i+1} \ll L_{i+1}$  and  $|L_{i+1} - L_i| < L_{i+1}/2$ , which are reasonable assumptions to make for MD simulations when the sampling interval is sufficiently small. We therefore obtain

$$u_{i+1}^{\text{TOR}} \equiv u_i^{\text{TOR}} + \left( \frac{L_{i+1}}{L_i} - 1 \right) w_i + \sigma_w R_{i+1}$$

in all practical cases.

By contrast, the LAT scheme evaluates to

$$\begin{aligned} u_{i+1}^{\text{LAT}} &= u_i^{\text{LAT}} + \left( \frac{L_{i+1}}{L_i} - 1 \right) w_i + \sigma_w R_{i+1} \\ &\quad - \left[ \frac{w_i - u_i^{\text{LAT}}}{L_i} + \frac{1}{2} \right] (L_{i+1} - L_i) \end{aligned} \quad (7)$$

when applied to the process of eq 5. Here, the last term can be further simplified via eq 3, giving

$$\begin{aligned} u_{i+1}^{\text{LAT}} &= u_i^{\text{LAT}} + \left( \frac{L_{i+1}}{L_i} - 1 \right) w_i + \sigma_w R_{i+1} - \left( \frac{L_{i+1}}{L_i} - 1 \right) \\ &\quad \times n_i L_i \\ &= u_i^{\text{LAT}} + \left( \frac{L_{i+1}}{L_i} - 1 \right) u_i^{\text{LAT}} + \sigma_w R_{i+1}. \end{aligned} \quad (8)$$

Comparing eqs 6 and 8, we find that the LAT scheme gives rise to a multiplicative noise term  $(L_{i+1}/L_i - 1) u_i^{\text{LAT}}$  that scales with the unwrapped coordinate. Its magnitude therefore grows without bounds as the particle diffuses away from the origin. The unbounded multiplicative noise in the LAT scheme causes pathological particle dynamics, which becomes apparent when

the LAT scheme is used to unwrap trajectories from Brownian dynamics (BD) and MD simulations, as demonstrated below in section 5.

**3.3. Consistent (Re)Wrapping Schemes.** Besides criticizing the undesired effect of intramolecular bond stretching, Kulke and Vermaas<sup>3</sup> further claimed that the TOR scheme cannot be reversible, because a subsequent wrapping of  $u^{\text{TOR}}$  using “conventional wrapping schemes” does not reproduce the wrapped trajectory  $w$ . While the authors did not explicitly specify which wrapping schemes they were referring to, we expect a lattice-view scheme, which in 1D reads

$$w_i^{\text{LAT}} = u_i^{\text{LAT}} - \left[ \frac{u_i^{\text{LAT}}}{L_i} + \frac{\alpha}{2} \right] L_i. \quad (9)$$

Here, the value of  $\alpha$  depends on the definition of the central unit cell. If it is defined by the interval  $[0, L_i]$  (as is the case, e.g., in GROMACS<sup>4</sup>) then  $\alpha = 0$ , whereas for cells fluctuating symmetrically around the origin, i.e.,  $[-L_i/2, L_i/2]$ , one has  $\alpha = 1$  (this is the convention that LAMMPS<sup>8</sup> and NAMD<sup>10</sup> adhere to). The scheme in eq 9 assumes that at each time integration step  $i$  the wrapped and unwrapped trajectories are identical up to an integer number of box lengths  $L_i$ , consistent with the lattice view of eq 3. Equation 9 should therefore be able to perfectly rewrap a trajectory generated by the LAT unwrapping scheme. In fact, substituting eq 7 into eq 9 with  $n = \lfloor (w_i^{\text{LAT}} - u_i^{\text{LAT}})/L_i + 1/2 \rfloor$  and  $\alpha = 1$  gives

$$w_{i+1}^{\text{LAT}} = \frac{L_{i+1}}{L_i} w_i^{\text{LAT}} + (n - n_i) L_i + \sigma_w R_{i+1} - \left[ \frac{w_i^{\text{LAT}}}{L_i} + (n - n_i) \frac{L_i}{L_{i+1}} + \frac{\sigma_w R_{i+1}}{L_{i+1}} + \frac{1}{2} \right] L_{i+1},$$

which coincides with eq 5 because

$$n = \left\lfloor \frac{w_i^{\text{LAT}} - u_i^{\text{LAT}}}{L_i} + \frac{1}{2} \right\rfloor = \left\lfloor n_i + \frac{1}{2} \right\rfloor \equiv n_i$$

must hold. In light of the fact that the TOR and LAT unwrapping schemes give different results, it is apparent that eq 9 cannot be used to correctly rewrap  $u^{\text{TOR}}$ .

To construct a (re)wrapping scheme consistent with the TOR unwrapping scheme, we backtrace the displacements  $u_{i+1}^{\text{TOR}} - u_i^{\text{TOR}}$  to reconstruct the wrapped trajectory in an iterative manner as follows:

$$w_0^{\text{TOR}} = u_0^{\text{TOR}},$$

$$w_{i+1}^{\text{TOR}} = w_i^{\text{TOR}} + (u_{i+1}^{\text{TOR}} - u_i^{\text{TOR}}) - \left[ \frac{w_i^{\text{TOR}} + (u_{i+1}^{\text{TOR}} - u_i^{\text{TOR}})}{L_{i+1}} + \frac{\alpha}{2} \right] L_{i+1}. \quad (10)$$

Whenever the trajectory  $w_i^{\text{TOR}} + (u_{i+1}^{\text{TOR}} - u_i^{\text{TOR}})$  crosses the periodic boundaries, it gets shifted back into the central box with the help of the last term. Substituting eq 6 into eq 10 gives rise to eq 5, as expected. Equations 9 and 10, and their relations to the LAT and TOR unwrapping schemes, are verified with the help of numerical data in section 5.

**3.4. Upper Bound for the Time Interval Between Sampled Structures in MD Simulations.** In section 3.2 we made the assumption that  $\sigma_w R_{i+1} \ll L_{i+1}$  must hold, and argued

for its validity for MD simulations at sufficiently small sampling intervals  $\Delta t$ . Very roughly, for a particle with mass  $m$  and diffusion coefficient  $D$  we require  $D\Delta t \ll \bar{L}^2$  and  $\Delta t^2 \ll \beta m \bar{L}^2$  in the regimes dominated by diffusion and inertia, respectively. Here,  $\beta^{-1} = k_B T$  denotes the thermal energy scale,  $k_B$  is the Boltzmann constant and  $T$  is the absolute temperature. To obtain a quantitative estimate of an upper bound for  $\Delta t$ , we consider the probability  $P$  for one or more of the  $N$  particles in the simulation box to travel a distance greater than  $\bar{L}/2$  within  $\Delta t$  along any of the three Cartesian coordinates, resulting in an incorrect unwrapping event in at least one of the  $t_{\text{total}}/\Delta t$  sampled frames, i.e.,

$$P = 1 - [1 - \Pr(X > \bar{L}/2)]^{3Nt_{\text{total}}/\Delta t}. \quad (11)$$

Here,  $t_{\text{total}}$  is the total simulation time and  $\Pr(X > \bar{L}/2)$  denotes the probability for a single particle in 1D to move either ballistically or diffusively in the time interval  $\Delta t$  by more than  $\bar{L}/2$  in either direction. We can estimate this probability as follows:

$$\Pr(X > \bar{L}/2) \approx 2 \int_{\bar{L}/2}^{\infty} dx \frac{e^{-x^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}} = \text{erfc}\left(\frac{\bar{L}}{2\sqrt{2\sigma^2}}\right),$$

where  $\text{erfc}(\cdot)$  denotes the complementary error function and

$$\sigma^2 = \begin{cases} 2D\Delta t, & \text{diffusive motion} \\ \Delta t^2/\beta m, & \text{ballistic motion} \end{cases}$$

depends on the dominating particle dynamics. For sufficiently small sampling intervals  $\Delta t$ , the probabilities  $P$  and  $\Pr(X > \bar{L}/2)$  are also small, which allows us to reduce eq 11 to  $P \approx 3N\Pr(X > \bar{L}/2) t_{\text{total}}/\Delta t$  and replace  $\text{erfc}(\cdot)$  with its asymptotic expansion for large arguments. Setting  $P = \varepsilon \ll 1$  results in the expression

$$\varepsilon \approx \frac{6Nt_{\text{total}}}{\bar{L}\Delta t} \sqrt{\frac{2\sigma^2}{\pi}} e^{-\bar{L}^2/8\sigma^2}, \quad (12)$$

which can be used to estimate how large the sampling interval  $\Delta t$  can be chosen for a specific MD simulation setup without skewing the resulting unwrapped trajectory due to particles crossing the periodic boundaries.

Solving eq 12 for  $\Delta t$  in the ballistic ( $\sigma^2 = \Delta t^2/\beta m$ ) and diffusive limit ( $\sigma^2 = 2D\Delta t$ ) gives the requirements

$$\Delta t_{\text{ballistic}} \leq \sqrt{\frac{\beta m \bar{L}^2}{8}} / \ln\left(\frac{6Nt_{\text{total}}}{\varepsilon \bar{L}} \sqrt{\frac{2}{\pi \beta m}}\right), \quad (13)$$

$$\Delta t_{\text{diffusive}} \leq -\frac{\bar{L}^2}{8D} / W_{-1}\left(-\left[\frac{\sqrt{\pi} \varepsilon \bar{L}^2}{24\sqrt{2} N D t_{\text{total}}}\right]^2\right), \quad (14)$$

respectively, where  $W_{-1}(\cdot)$  denotes the lower branch of the Lambert  $W$  function, which can be expanded for small arguments to give  $W_{-1}(z \rightarrow 0^-) \sim -\ln(-1/z) - \ln(\ln(-1/z)) - \ln(\ln(-1/z)/\ln(-1/z))$ .<sup>11</sup> For  $\varepsilon = 10^{-2}$  and  $N \approx 33.3 \text{ nm}^{-3} \bar{L}^3$  TIP3P water molecules with  $D \approx 6 \text{ nm}^2 \text{ ns}^{-1}$ <sup>12</sup> and  $m \approx 18 \text{ g mol}^{-1}$  at  $T = 300 \text{ K}$ , eqs 13 and 14 evaluate to  $\Delta t_{\text{ballistic}} \approx 0.48 \text{ ps}$ ,  $0.94 \text{ ps}$ ,  $1.4 \text{ ps}$  and  $\Delta t_{\text{diffusive}} \approx 2.9 \text{ ps}$ ,  $11 \text{ ps}$ ,  $25 \text{ ps}$  for cubic simulation boxes with edge lengths  $\bar{L} = 2.5 \text{ nm}$ ,  $5 \text{ nm}$ , and  $7.5 \text{ nm}$ , respectively. For water at ambient conditions,<sup>2</sup> one can safely use the  $\Delta t_{\text{diffusive}}$  bound for times  $\Delta t \geq 1 \text{ ps}$ .



## 4. METHODS

### 4.1. MD Simulation of TIP4P-D Water with GROMACS.

We made use of a 1  $\mu$ s constant-pressure simulation of 515 TIP4P-D water molecules<sup>13</sup> in a cubic box with an average edge length of  $\bar{L} \approx 2.5$  nm, which was previously reported on in ref 2. The simulation was run using GROMACS 2018.6<sup>4</sup> with a 2 fs integration time step, and particle-mesh Ewald electrostatics<sup>14</sup> with a 1.2 nm real-space cutoff. The SETTLE algorithm was used to keep water molecules rigid.<sup>15</sup> The production run commenced after a 100 ps initial equilibration at constant volume and a subsequent 5 ns equilibration run at constant pressure. Temperature and pressure were maintained at 300 K and 1 bar throughout the entire simulation using the velocity-rescaling thermostat<sup>16</sup> ( $\tau_T = 1$  ps) and the Parrinello–Rahman barostat<sup>17</sup> ( $\tau_p = 5$  ps), respectively. Particle coordinates were sampled every  $\Delta t = 1$  ps. For the sake of comparison, we also considered a 1  $\mu$ s constant-volume simulation of the system described above, which was equilibrated and run in an identical manner.<sup>2</sup>

**4.2. MD Simulation of SPC/E Water with LAMMPS.** We generated a set of wrapped and unwrapped trajectories of 511 SPC/E water molecules<sup>18</sup> at ambient conditions using the LAMMPS package stable release from September 29, 2021 (update 3).<sup>8,19</sup> The simulation was performed at constant pressure in a cubic box with an average edge length of  $\bar{L} \approx 2.5$  nm. The SHAKE algorithm<sup>20</sup> was used to constrain the intramolecular bonds and angles at an accuracy tolerance of  $10^{-4}$ . The particle-particle particle-mesh solver<sup>21</sup> with a relative force error accuracy of  $10^{-4}$  was used to compute long-range Coulombic interactions, where the cutoff distance in real space was set to 9.8 Å. Equilibration consisted of a 15 ns run in the NVT ensemble, followed by a 20 ns run in the NPT ensemble. Temperature and pressure were maintained at 300 K and 1 bar using the Nosé–Hoover thermostat and barostat<sup>22,23</sup> with damping coefficients of 100 and 1000 fs, respectively. The 1  $\mu$ s production run in the NPT ensemble was performed using the same thermostat and barostat coefficients, and a 1 fs integration time step. Particle coordinates of the wrapped and unwrapped trajectory were sampled every  $\Delta t = 1$  ps via the dump command.

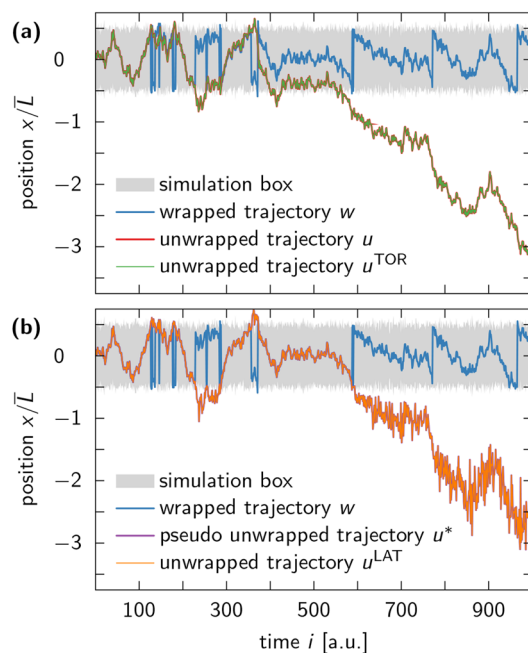
**4.3. MD Simulation of TIP3P Water with NAMD.** We generated an unwrapped trajectory of 826 TIP3P water molecules<sup>24</sup> at ambient conditions in a cubic periodic box with  $\bar{L} \approx 2.9$  nm, using NAMD version 3.<sup>10</sup> A time step of 2 fs was used. Temperature was maintained at 300 K using underdamped Langevin dynamics with a damping time of 1 ps. Pressure was set to 1 bar using the Nosé–Hoover Langevin piston method as implemented in NAMD,<sup>25</sup> with a piston period of 200 fs and a decay time of 100 fs. Water molecules were kept rigid using the SETTLE algorithm.<sup>15</sup> Long-range electrostatic interactions were computed using the Particle-Mesh Ewald method, with a 12 Å cutoff for the real-space part. The same cutoff was applied to Lennard-Jones potentials, with force-switching for a continuous decay of the force to zero. Particle coordinates were sampled every  $\Delta t = 1$  ps for 900 ns in total, this slightly shorter duration being the result of numerical instability (see further section 5.3). To obtain a corresponding wrapped trajectory, we chose to wrap the NAMD output trajectory using eq 9.

## 5. RESULTS AND DISCUSSION

**5.1. Brownian Dynamics Simulations.** To verify the analytic predictions of section 3, we evaluated eqs 5, 6, and 8 in

an iterative fashion to generate a wrapped trajectory  $w$ , and two unwrapped trajectories  $u$  and  $u^*$ . The wrapped trajectory was unwrapped using the TOR and LAT schemes (eqs 2 and 4, respectively), and the resulting unwrapped trajectories were compared to the corresponding  $u$  and  $u^*$  realizations. We used random initial positions  $w_0 = u_0 = u_0^* = \bar{L}(R' - 1/2)$  with  $R' \sim \mathcal{U}_{[0,1]}$  uniformly distributed on the interval  $[0,1]$ , and fixed parameter values of  $\sigma_L = 0.1\bar{L}$  and  $\sigma_w = 0.05\bar{L}$ .

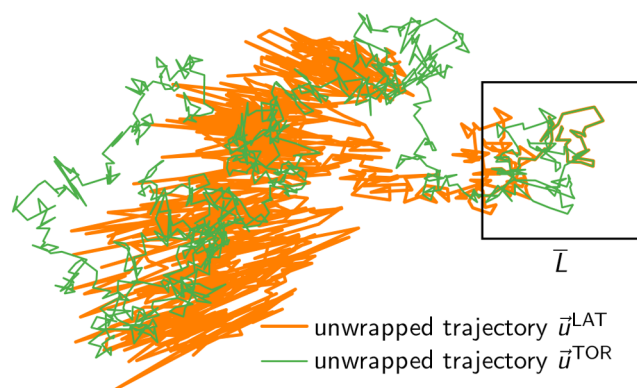
Figure 2 displays a representative set of trajectories that were generated as described above. In accordance with our



**Figure 2.** Comparison of the TOR and LAT unwrapping schemes in 1D. (a) TOR unwrapping of trajectory for 1D Gaussian model. While the wrapped trajectory  $w$  (blue line, eq 5) is confined to the simulation box (gray shaded area), its unwrapped partner trajectory  $u$  (red line, eq 6) can traverse arbitrarily far from their common initial position. The TOR unwrapping scheme (eq 2), when applied to  $w$ , produces a trajectory  $u^{\text{TOR}}$  (green line), which completely overlaps with  $u$ . Note that the unwrapped trajectories  $u$  and  $u^{\text{TOR}}$  are not “on lattice” in NPT simulations. As a result, they may not coincide with the wrapped trajectory  $w$  in revisits to the central simulation box, as seen around time 500. (b) LAT unwrapping of the same trajectory as in (a). The unwrapped trajectory  $u^{\text{LAT}}$  (orange line) generated by the LAT unwrapping scheme (eq 4) coincides with the pseudo unwrapped trajectory  $u^*$  (purple line, eq 8) and exhibits the same exaggerated fluctuations away from the central box.

predictions in section 3.2, the unwrapped trajectory  $u^{\text{LAT}}$  associated with the LAT scheme exhibits the same position-dependent fluctuations that can be found in  $u^*$  (eq 8), which increase with the distance to the origin in stark contrast to the dynamics of the wrapped trajectory. Meanwhile, the unwrapped trajectory  $u^{\text{TOR}}$  generated by the TOR scheme shows moderate fluctuations and completely overlaps with  $u$ , as expected. A visual comparison of trajectory segments between two boundary-crossing events demonstrates that  $u^{\text{TOR}}$  perfectly captures the trends observed in  $w$ . The same cannot be said about  $u^{\text{LAT}}$ .

The nondiffusivity of  $u^{\text{LAT}}$  is even more pronounced in higher dimensions, as illustrated in Figure 3, where we combine two 1D trajectories of the Gaussian model,  $w_x$  and  $w_y$ , to construct a two-



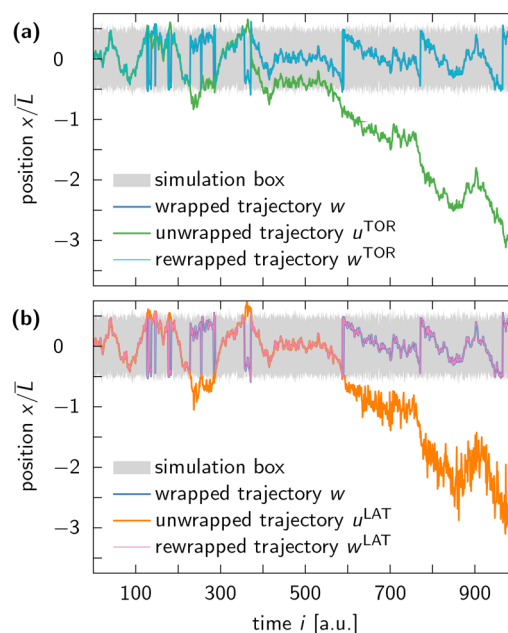
**Figure 3.** Comparison of the TOR and LAT unwrapping schemes in 2D. A wrapped trajectory  $\vec{w}$  of the Gaussian model (not shown) was unwrapped using the TOR and LAT schemes, which resulted in the unwrapped trajectories  $\vec{u}^{\text{TOR}}$  (green line) and  $\vec{u}^{\text{LAT}}$  (orange line), respectively. While  $\vec{u}^{\text{TOR}}$  is visually indistinguishable from an ordinary diffusive trajectory anywhere in the plane,  $\vec{u}^{\text{LAT}}$  is strongly affected by box fluctuations after leaving the central simulation box (average size shown as black square). Note that the apparent noise in  $\vec{u}^{\text{LAT}}$  grows with the distance from the central box and becomes anisotropic, with position fluctuations emanating in a star-like fashion from the origin.

dimensional (2D) wrapped trajectory  $\vec{w} = (w_x, w_y)^T$ . The unwrapped trajectories  $\vec{u}^{\text{TOR}}$  and  $\vec{u}^{\text{LAT}}$  were generated by unwrapping each component of  $\vec{w}$  separately using the TOR and LAT schemes, respectively. As shown, the apparent noise in the  $\vec{u}^{\text{LAT}}$  trajectory not only grows with distance from the central simulation box but also becomes anisotropic. [Supporting Movie S1](#) visualizes the evolution of the trajectories and the fluctuations of the simulation box.

Finally, we assessed the ability of eqs 9 and 10 to reverse the operations of the LAT and TOR unwrapping schemes, respectively. In [Figure 4](#), we demonstrate that eq 10 faithfully reproduces the wrapped 1D trajectory  $w$  when applied to  $u^{\text{TOR}}$ . Similarly, we find that eq 9 perfectly rewraps  $u^{\text{LAT}}$  back into the simulation box. It is therefore unsurprising that Kulke and Vermaas<sup>3</sup> only found their own unwrapping scheme to be reversible with respect to “conventional wrapping schemes” (such as eq 9): the unwrapped trajectories  $u^{\text{TOR}}$  and  $u^{\text{LAT}}$  are different and therefore require different wrapping schemes to be correctly rewrapped into the simulation box.

**5.2. GROMACS Simulations.** In our BD simulations, we could freely choose the amplitude  $\sigma_L$  of the box fluctuations to highlight the difference between the two unwrapping schemes. In MD simulations, however, box fluctuations for aqueous systems at ambient conditions are generally well below one percent of the average edge length, so the amplified fluctuations in  $u^{\text{LAT}}$  are much more subtle. To test whether we can identify considerable differences between the TOR and LAT schemes in MD simulations, we analyzed the wrapped trajectories of TIP4P-D water in a small, fluctuating, cubic box, as reported previously in ref 2 (see [section 4.1](#) for technical details).

In [Figure 5](#), we plot the  $y$ -component of the trajectory of an oxygen atom in a water molecule that managed to diffuse more than 30 box edge lengths away from the central simulation box. At first glance, the unwrapped trajectories produced by the TOR and LAT schemes may seem identical, but when we zoom in on the last few nanoseconds of the trajectory, we find  $u^{\text{LAT}}$  to have the same exaggerated fluctuations as observed in our BD



**Figure 4.** Rewrapping trajectories using appropriate wrapping schemes. (a) The unwrapped trajectory  $u^{\text{TOR}}$  (green line), generated by the TOR scheme, can be perfectly rewrapped inside the simulation box (gray shaded area) using eq 10 with  $\alpha = 1$ , as seen by the complete overlap of  $w^{\text{TOR}}$  (cyan line) with the original wrapped trajectory  $w$  (blue line). (b) Similar results can be achieved for  $u^{\text{LAT}}$  (orange line), generated by the LAT scheme, if it is rewrapped using eq 9 with  $\alpha = 1$ . This gives rise to the trajectory  $w^{\text{LAT}}$  (pink line).

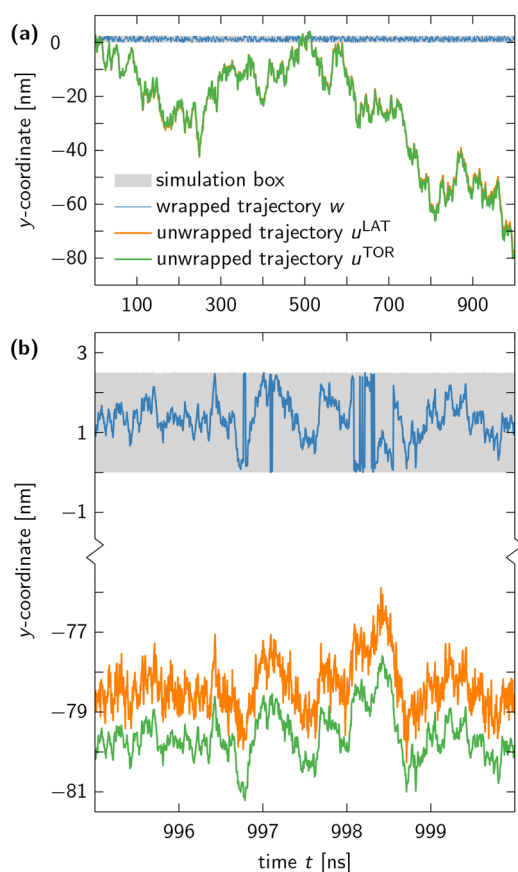
simulations. By contrast,  $u^{\text{TOR}}$  visually reproduces the features of the wrapped trajectory  $w$ .

To quantify the effect that box fluctuations have on unwrapped trajectories, we analyzed the diffusive behavior observed in different trajectory segments. We made use of a maximum likelihood estimator<sup>26</sup> (MLE) for the diffusion coefficient  $D$ , which accounts for the fact that a  $d$ -dimensional diffusive process  $\vec{X}(t)$  can be corrupted by static noise and dynamic motion blur, resulting in the following mean squared displacement (MSD):

$$\text{MSD}(\tau) = \langle |\vec{X}(\tau) - \vec{X}(0)|^2 \rangle = a^2 d + 2dD(\tau - 2B). \quad (15)$$

Here,  $\tau$  denotes the lag time. For MD simulations, the motion blur coefficient  $B$  is zero and the vertical intercept  $a^2$  accounts for nondiffusive dynamics at short times.<sup>27</sup> Note that the MLE does not rely explicitly on estimates for the MSD, but instead exploits the statistics of the increments  $\vec{X}(t_{i+1}) - \vec{X}(t_i)$ . We segmented our unwrapped oxygen trajectories for TIP4P-D water into 1 ns blocks and extracted for each block (with index  $i$ ) estimates for the static noise  $a_i^2$  and the diffusion coefficient  $D_i$ . [Figure 6](#) presents our results for the TOR and LAT schemes. Unsurprisingly, the TOR scheme gives consistent parameter estimates for all block indices  $i$ , whereas the estimates for the LAT scheme vary strongly with  $i$  and tend toward larger values at later times in the trajectory. This behavior is to be expected as the diffusive spread of the water molecules moves them further away from the central simulation box, where artifacts become more pronounced in trajectories associated with the LAT unwrapping scheme.

It should be noted that our results imply a significant difference between global diffusion coefficient estimates



**Figure 5.** Trajectory of an oxygen atom of a TIP4P-D water molecule along a single coordinate axis. (a) The wrapped trajectory  $w$  (blue line) is unwrapped via the TOR and LAT schemes, resulting in  $u^{\text{TOR}}$  (green line) and  $u^{\text{LAT}}$  (orange line), respectively. The two unwrapped trajectories seem almost identical, because box fluctuations in MD simulations of water at ambient conditions are small compared to the dimensions of the simulation box (gray shaded area). (b) However, a zoom-in on the last 5 ns of the trajectory reveals that  $u^{\text{LAT}}$  exhibits larger fluctuations between subsequent time frames than  $u^{\text{TOR}}$  and  $w$ . The enhanced noise in  $u^{\text{LAT}}$  is indicative of the unbounded multiplicative noise associated with the LAT unwrapping scheme.

obtained for the TOR scheme and the LAT scheme with a  $\Delta t = 1$  ps time step. Consistent with the analysis of Figure 6, we find the global mean estimates  $\bar{D}_{\text{TOR}} = 2.0602(2) \text{ nm}^2 \text{ ns}^{-1}$  and  $\bar{D}_{\text{LAT}} = 4.64(3) \text{ nm}^2 \text{ ns}^{-1}$  by applying the MLE to the full  $1 \mu\text{s}$  trajectory of each oxygen atom and then average over all water molecules, without correcting for finite-size effects.<sup>12,28</sup> By comparison, we obtained the global mean estimate  $\bar{D} = 2.0517(2) \text{ nm}^2 \text{ ns}^{-1}$  for the corresponding NVT simulation (see section 4.1 for technical details). The standard errors of  $\bar{D}_{\text{TOR}}$ ,  $\bar{D}_{\text{LAT}}$ , and  $\bar{D}$  (in parentheses) were estimated by assuming that the diffusion processes of individual water molecules were uncorrelated. Importantly, we expect the discrepancy between  $\bar{D}_{\text{TOR}}$  and  $\bar{D}_{\text{LAT}}$  to grow if the MD simulations are extended beyond the  $1 \mu\text{s}$  used here, because  $D$  estimates in Figure 6 obtained with the MLE for the LAT unwrapped trajectories slowly increase to ever larger values as the traced particles move further away from the central box (see Figure 6).

**5.3. LAMMPS and NAMD Simulations.** The GROMACS simulation software package exclusively generates wrapped trajectories. These trajectories are typically unwrapped in a postprocessing step using built-in tools like `trjconv` or third-

party software, such as the PBCTools<sup>29</sup> and `qwrap`<sup>7</sup> plugins for VMD.<sup>30</sup> Other MD simulation codes write out unwrapped trajectories directly, either by default or via user-specified settings, but this raises the question which unwrapping scheme these trajectories correspond to. We therefore analyzed simulation trajectories for SPC/E and TIP3P water generated via the software packages LAMMPS and NAMD, respectively (see sections 4.2 and 4.3 for technical details). NAMD does not, in general, wrap the particle coordinates throughout the simulation, except when writing coordinates to disk, and then only when instructed to do so through the user options `wrapAll` or `wrapWater`. LAMMPS, by contrast, allows the user to specify whether the wrapped coordinates, unwrapped coordinates, or both should be written out.

We segmented the unwrapped trajectories of oxygen atoms generated by LAMMPS and NAMD into 1 ns blocks and analyzed the diffusive dynamics of every block separately, as detailed in section 5.2. This was also done to the corresponding wrapped partner trajectories, after unwrapping them via the TOR unwrapping scheme. The resulting diffusion coefficient estimates as functions of the time window are shown in Figure 7.

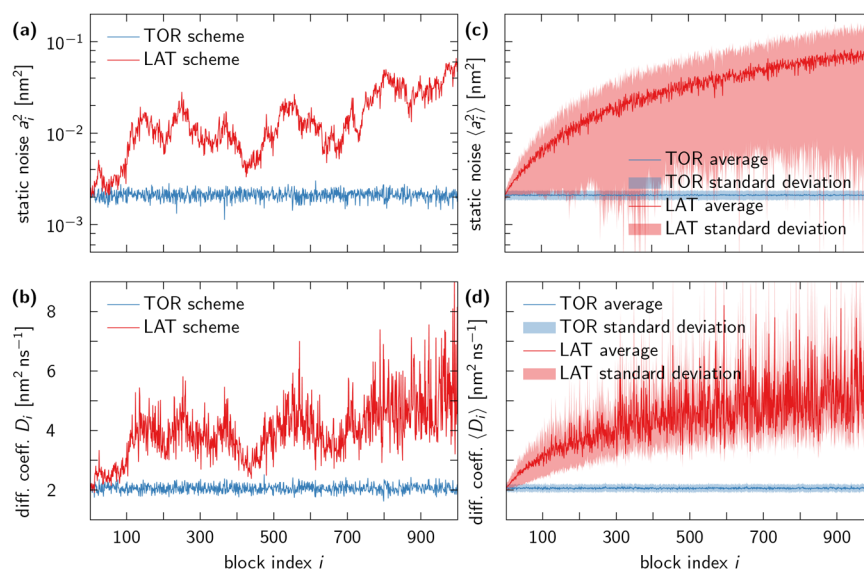
We find that the diffusion coefficients calculated directly from the unwrapped trajectories steadily diverge to ever larger values as a function of time. By contrast, the wrapped trajectories, which were unwrapped using the TOR scheme, produce robust diffusion coefficient estimates. The behavior in Figure 7 mirrors that seen in Figures 6(c) and 6(d), and is fully consistent with LAMMPS and NAMD producing unwrapped NPT trajectories that are “on lattice” and thus not suitable for the estimation of diffusion coefficients. Indeed, the global mean diffusion coefficients obtained by analyzing the entire unwrapped trajectories produced directly (i.e., without division into blocks) by LAMMPS and NAMD are inconsistent with the results for early and late blocks. The global mean is  $\bar{D}_{\text{LAMMPS}} = 3.75(1) \text{ nm}^2 \text{ ns}^{-1}$  for SPC/E water in the LAMMPS simulation, whereas early and late blocks give values around  $2.7 \text{ nm}^2 \text{ ns}^{-1}$  and  $4 \text{ nm}^2 \text{ ns}^{-1}$ , respectively [see Figure 7(a)]. For TIP3P water in the NAMD simulations, the global mean is  $\bar{D}_{\text{NAMD}} = 6.50(2) \text{ nm}^2 \text{ ns}^{-1}$ , with early and late blocks around  $5 \text{ nm}^2 \text{ ns}^{-1}$  and  $7 \text{ nm}^2 \text{ ns}^{-1}$ , respectively [see Figure 7(b)]. By contrast, the global mean  $\bar{D}_{\text{TOR}} = 2.7078(2) \text{ nm}^2 \text{ ns}^{-1}$  obtained after TOR unwrapping of the wrapped LAMMPS trajectory is consistent with the respective block estimates, as is  $\bar{D}_{\text{TOR}} = 5.0132(3) \text{ nm}^2 \text{ ns}^{-1}$  obtained for the NAMD trajectory. Note that we did not correct for system-size effects on the self-diffusion coefficients.<sup>12,28</sup>

The NAMD simulation became numerically unstable as particle coordinates and their barostat-induced fluctuations became large. While such a small box with less than 1000 water molecules can be seen as an extreme example, this phenomenon highlights a benefit of propagating wrapped coordinates internally during the simulation, which is to make the best use of limited floating-point precision, especially in mixed-precision GPU software. In NAMD this behavior can be approached by enabling the `wrapAll` or `wrapWater` options, resulting in all coordinates or specifically water molecules being wrapped at every restart.

#### 5.4. Are Published Diffusion Coefficients from NPT Simulations Compromised by Incorrect Unwrapping?

To assess the possible impact of LAT unwrapping on diffusion coefficients obtained from NPT simulations in earlier studies, it is important to recognize that the vast majority of  $D$  estimates were calculated from ordinary least-squares (OLS) fits of straight lines to the MSD. The fit region is usually chosen “by





**Figure 6.** Diffusion coefficient estimates are robust for TOR unwrapping but compromised by LAT unwrapping. Shown are results for the static noise and diffusion coefficient estimates obtained using the MLE from trajectories of oxygen atoms in TIP4P-D water saved at a time interval of  $\Delta t = 1$  ps and divided into 1000 blocks  $i$  of 1 ns each. (a) Static noise amplitude  $a_i$  and (b) diffusion coefficient  $D_i$  estimated for each block  $i$  of the unwrapped trajectory of a single water molecule using the TOR (blue lines) and LAT schemes (red lines). (c) Average static noise amplitude  $a_i$  and (d) diffusion coefficient  $D_i$  estimates over all water molecules using the TOR (blue) and LAT schemes (red). In (c, d), averages are shown as solid lines and standard deviations as shaded areas.

eye” to cover a window in which the MSD grows more or less linearly with time. Considering the fact that, to a rough approximation, LAT trajectories make ever larger jumps about TOR trajectories, it is conceivable that simple  $D$  estimates obtained from OLS fits to the MSD are less impacted by LAT unwrapping than the more sophisticated MLE used in the analysis above.

In Figure 8, we present the MSD of LAT and TOR trajectories (averaged over all trajectories) at different lag times  $\tau = m\Delta t$  for the TIP4P-D water data analyzed in section 5.2. The single-trajectory MSD values were estimated as

$$\text{MSD}(m\Delta t) = \sum_{n=0}^{N-m} \frac{|\vec{X}((n+m)\Delta t) - \vec{X}(n\Delta t)|^2}{N-m+1},$$

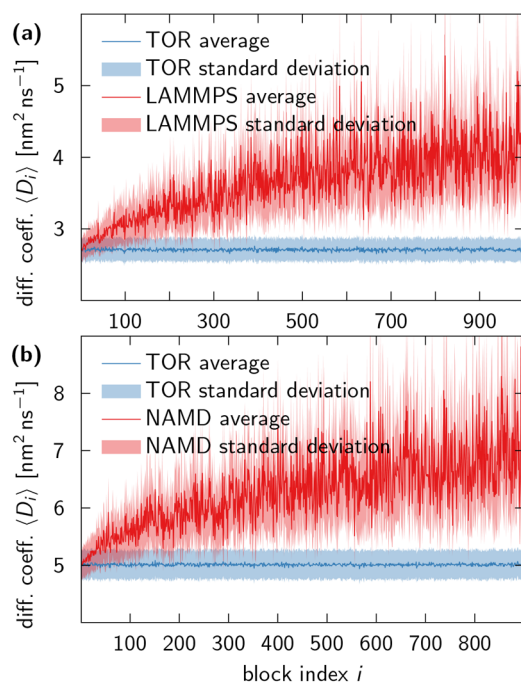
with  $N = 10^6$  data points for a time step of  $\Delta t = 1$  ps. The MSD obtained from the LAT unwrapped trajectories oscillates at short lag times and then approaches a straight line. Halving the simulation time to 500 ns reduces the magnitude of the oscillations at short times, consistent with the observation that the amplitude of oscillations in the LAT unwrapped trajectories tends to increase with their duration (Figure 3). At long lag times, however, the slope of the MSD from LAT unwrapping approaches that obtained from TOR unwrapping. This is confirmed by OLS fits, where we found good agreement between diffusion coefficient estimates for the TOR and LAT data. These findings explain why no significant difference was found between  $D_{\text{TOR}}$  and  $D_{\text{LAT}}$  in ref 3. For completeness, we also analyzed the unwrapped trajectories using a generalized least-squares estimator<sup>27</sup> (GLS) and a covariance-based estimator<sup>31</sup> (CVE). The resulting estimates of the diffusion coefficient of TIP4P-D water (without correcting for system-size effects<sup>12,28</sup>) are listed in Table 1. Note that the slight discrepancy between the MSD for the TOR unwrapped trajectories at long lag times  $\tau$  and eq 15, evaluated using the estimates of MLE, GLS, and CVE (see Figure 8), results from the fact that the

dynamics of TIP4P-D water is not perfectly diffusive at  $\Delta t = 1$  ps. This was discovered and extensively discussed in ref 27, where a time step of  $\Delta t = 10$  ps was identified as optimal, in the sense that it is short enough to minimize statistical errors but long enough to ensure diffusive dynamics. In fact, if the LAT and TOR trajectories, which were unwrapped at  $\Delta t = 1$  ps, are subsampled at  $\Delta t = 10$  ps and then reanalyzed, one gets fairly similar diffusion coefficients for both unwrapping schemes and all estimators considered here (see Table 1). The TOR-trajectory estimates of MLE, GLS, and CVE all lie within their respective statistical uncertainties.

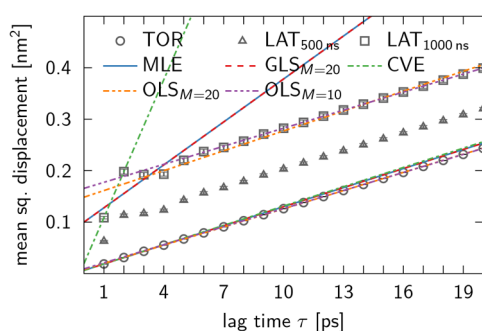
As a further consistency check, we calculated the TIP4P-D water diffusion coefficient also in the NVT ensemble, where unwrapping is unambiguous and the TOR and LAT schemes are equivalent. The NVT simulations were performed for a system of the same size and temperature, with a volume calibrated to produce a pressure close to that of the NPT simulations.<sup>2</sup> We found that the diffusion coefficient estimates obtained with the MLE for trajectories in the NVT ensemble sampled at time steps of  $\Delta t = 1$  and 10 ps are both in very close agreement with the respective estimates in the NPT ensemble obtained with TOR unwrapping (Table 1). This consistency indicates that the box-size fluctuations caused by the barostat did not noticeably impact the statistical properties of the molecular diffusion process for NPT trajectories unwrapped with the TOR scheme.

Reassuringly, the results in Figure 8 and Table 1 suggest that earlier diffusion coefficients may actually be less impacted by LAT unwrapping than the results of Figures 6 and 7 suggest, for the simple reason that in the MD community, straight-line fits to the MSD are still commonly used to estimate  $D$ . Moreover, early simulations tended to be comparably short, so that particles did not diffuse far from the simulation box at the origin. Going forward, however, we recommend to take advantage of the statistically more efficient MLE<sup>26</sup> or GLS<sup>27</sup> estimators applied to TOR-unwrapped trajectories, which preserve the diffusive characteristics also at long times, unlike the LAT trajectories.





**Figure 7.** Diffusion coefficient estimates for default-unwrapped NAMD trajectories and automatically unwrapped LAMMPS trajectories are compromised. (a) LAMMPS simulation of SPC/E water. Diffusion coefficients were estimated separately for 1000 consecutive blocks  $i$ , each 1 ns long, of a continuous trajectory created by writing out “unwrapped” coordinates (red). A corresponding wrapped trajectory, also written out by LAMMPS, was unwrapped using the TOR scheme, and analyzed analogous to the automatically unwrapped LAMMPS trajectory to produce the TOR estimates of the diffusion coefficient (blue). (b) NAMD simulation of TIP3P water. The data analysis procedure was the same as used for the LAMMPS trajectories, except that the wrapped counterpart of the NAMD trajectory was generated using eq 9. Due to the fact that the NAMD trajectory was 100 ns shorter than the LAMMPS trajectory, it was split into 900 blocks. Averages are shown as solid lines and standard deviations as shaded areas.



**Figure 8.** LAT unwrapping compromises the MSD of TIP4P-D water trajectories at short lag times  $\tau$ . While the MSD of TOR unwrapped trajectories (circles) is linear at all lag times, its counterpart for LAT trajectories (triangles for the first 500 ns of the simulation and squares for the full simulation time) displays oscillations for small  $\tau$ , which grow in size with the length of the trajectories. The corresponding error bars are smaller than the markers. Lines represent eq 15 with values of  $a^2$  and  $D$  that were calculated using various estimators (see legend) for a time step of  $\Delta t = 1$  ps. The associated mean diffusion coefficients  $\bar{D}_{\text{TOR}}$  and  $\bar{D}_{\text{LAT}}$  are listed in Table 1.

Combining powerful estimators with TOR-unwrapped trajectories avoids systematic errors and minimizes statistical errors.

**5.5. Correctly Applying the TOR Scheme to Bonded Atoms.** Now that we have established the correctness and consistency of the TOR unwrapping scheme for single particles, we next address possible issues that can arise due to the fact that the scheme does not adhere to the lattice view and therefore cannot preserve distances between particles. This becomes a problem, for example, when the TOR scheme is applied naively to bonded particles.

Kulke and Vermaas<sup>3</sup> observed that bond lengths only got distorted when the TOR scheme was applied to trajectories generated by the GROMACS software package, whereas NAMD trajectories seemed unaffected. The reason for this discrepancy is the fact that NAMD by default treats molecules as “whole” when writing out data, i.e., the software does not break up molecules that sit on the periodic boundary. A preprocessing step in the analysis of GROMACS data, to make molecules whole prior to unwrapping, will therefore remedy the seeming shortcoming of the TOR scheme observed in ref 3.

Irrespective of the simulation software behind the data to be analyzed, we recommend the following order of operations when unwrapping MD simulation data of molecules:

1. In each frame of the trajectory, make the molecule “whole,” i.e., starting from a chosen reference atom of the molecule, ensure that all covalent bonds correspond to their minimal distance over the periodic images.
2. Calculate the center of mass of the “whole” molecule and, in case the resulting coordinate is located outside of the simulation box, perform a wrapping operation. This generates a wrapped trajectory of the center-of-mass coordinate of the molecule.
3. Unwrap the trajectory of the center-of-mass coordinate using the TOR unwrapping scheme.
4. If needed, the molecule can be reconstructed along the unwrapped center-of-mass trajectory by using the positions of the atoms relative to the center of mass of the “whole” molecule in each frame.

Note that the calculation of the center-of-mass coordinate in step 2 can be avoided by using instead the position of a specific atom as reference, say the oxygen atom of a water molecule. Also note that the estimation of translational diffusion coefficients only requires the tracking of the center of mass or any chosen reference atom.

**5.6. Pair Diffusion.** Finally, one might speculate whether the nonpreservation of distances in the TOR unwrapping scheme affects other observables that rely on unwrapping, such as pair diffusion coefficients. According to theory, the distance vector  $\vec{X} - \vec{Y}$  between two independent diffusion processes,  $\vec{X}(t)$  and  $\vec{Y}(t)$ , is also diffusive with the following diffusion coefficient:

$$D^{\text{pair}} = D^{\vec{X}} + D^{\vec{Y}}. \quad (16)$$

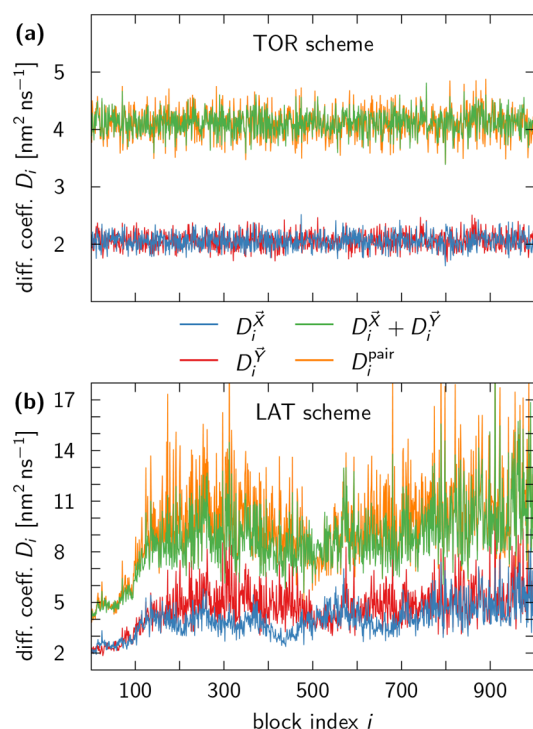
Here,  $D^{\vec{Z}}$  denotes the diffusion coefficient of the process  $\vec{Z}(t)$ .

To test whether pair diffusion is preserved for the TOR scheme, the LAT scheme, or both, we considered two randomly selected TIP4P-D water molecules from the GROMACS MD simulation and analyzed the diffusive behavior of  $\vec{X}$ ,  $\vec{Y}$ , and  $\vec{X} - \vec{Y}$ , as described in section 5.2. Figure 9 demonstrates that both unwrapping schemes essentially satisfy eq 16, but only the TOR scheme gives consistent results for all trajectory segments. As for single-particle diffusion, the pair diffusion coefficient obtained by LAT unwrapping tends to grow with time. It is therefore clear that LAT unwrapping and the associated lattice view of the

Table 1. Mean Diffusion Coefficients of TIP4P-D Water Calculated Using Various Estimators<sup>a</sup>

estimator	$m_{\text{init}}$	$m_{\text{end}}$	$\Delta t = 1$ ps time step		$\Delta t = 10$ ps time step	
			$\bar{D}_{\text{TOR}}$	$\bar{D}_{\text{LAT}}$	$\bar{D}_{\text{TOR}}$	$\bar{D}_{\text{LAT}}$
MLE <sup>26</sup>	1	20	2.0602(2)	4.64(3)	1.9523(4)	1.9442(5)
GLS <sup>27</sup>			2.0602(2)	4.64(3)	1.9523(4)	1.9442(5)
CVE <sup>31</sup>			2.0838(2)	14.9(4)	1.9526(4)	1.9381(7)
OLS <sub>M=20</sub>	1	20	1.9646(3)	2.147(5)	1.9475(9)	1.9465(9)
OLS <sub>M=10</sub>	11	20	1.9518(4)	1.9487(4)	1.947(1)	1.948(1)
MLE			$\bar{D}_{\text{NVT}} = 2.0517(2)$		$\bar{D}_{\text{NVT}} = 1.9507(4)$	

<sup>a</sup>The GLS and OLS estimators were evaluated for  $M = m_{\text{end}} - m_{\text{init}} + 1$  MSD values at lag times  $\tau = m \Delta t$  with  $m \in [m_{\text{init}}, m_{\text{end}}]$ . All diffusion coefficients are uncorrected for finite-size effects and have units of  $\text{nm}^2 \text{ns}^{-1}$ . Uncertainties in the last significant digit are listed in parentheses and correspond to standard errors of the mean.



**Figure 9.** Pair diffusion coefficient estimates for the TOR and LAT scheme both satisfy the additivity relation (eq 16) but are only robust for the TOR scheme. (a) Analogous to Figure 6(a), we analyzed 1 ns segments of two unwrapped trajectories generated by the TOR scheme and extracted the corresponding diffusion coefficients  $D_i$  (blue and red lines). The pair diffusion coefficient  $D_i^{\text{pair}}$  (orange line) agrees well with  $D_i^x + D_i^y$  (green line), but numerical discrepancies are due to the nonlinearity of our MLE. (b) Same as in panel (a) for trajectories unwrapped using the LAT scheme.

PBCs are not beneficial for the estimate of pair diffusion coefficients in constant-pressure NPT simulations.

A special case of restricted pair diffusion is the end-to-end relaxation of long (bio)polymers. Any motions that extend over the scale of the box will require care in unwrapping of PBCs. For polymer end-to-end relaxations and similar properties, one should first make the molecule “whole” across PBCs using the instantaneous box size and then calculate the relaxation dynamics of, say, the end-to-end distance.

## 6. CONCLUSIONS

Unwrapping trajectories of constant-pressure MD simulations is a nontrivial task. One school of thought takes a toroidal view of the PBCs to construct an unwrapped trajectory by adding up

minimal displacement vectors at each time step.<sup>2</sup> Another<sup>3,7</sup> takes a lattice view of the PBCs and traces the trajectory through the fluctuating lattice of image particles. As a consequence of the fluctuations in box size and shape, and the associated fluctuations in the lattice parameters, the two approaches produce different unwrapped trajectories. We have shown here that the toroidal approach embodied in the TOR algorithm<sup>2</sup> sacrifices the preservation of interparticle distances to preserve the statistical properties of the wrapped trajectory. In particular, a trajectory created by a diffusion process retains its diffusive character. By contrast, the lattice view taken in the LAT algorithm<sup>3</sup> preserves distances between particles, but distorts the statistical properties of local dynamics and thus destroys the diffusive character of a wrapped diffusion trajectory.

Our analytic calculations (section 3.2), as well as our results from BD and MD simulations (sections 5.1 and 5.2), demonstrate that the lattice-preserving LAT unwrapping scheme amplifies position-dependent fluctuations that arise in wrapped trajectories of constant-pressure simulations due to barostat position rescaling. Meanwhile, the TOR unwrapping scheme manages to preserve the diffusive character of the wrapped trajectories. These observations are further confirmed by our diffusion analysis of unwrapped MD trajectories, where different segments of TOR trajectories give consistent diffusion coefficient estimates, whereas the diffusive dynamics at the beginning and end of LAT trajectories differ greatly (see Figure 6 for water at ambient conditions).

A surprising conclusion is that the “unwrapped” trajectories written out by MD simulation software like NAMD and LAMMPS for NPT simulations should not be used to calculate diffusion coefficients (section 5.3). The reason is that these trajectories are “on lattice,” i.e., the corresponding wrapped positions are obtained by eq 9. Therefore, as the particles diffuse away from the reference box at the origin, they increasingly pick up the multiplicative noise resulting from box rescaling, as visualized in Figures 2 and 3. To avoid the resulting artifacts (see, e.g., Figure 7) and to obtain accurate diffusion coefficients, the output trajectories should first be wrapped “on lattice” via eq 9 and then unwrapped “off lattice” using the TOR scheme (eq 2).

In light of the fact that particle dynamics in constant-pressure simulations is always affected by a (bounded) barostat-induced multiplicative noise, even in the wrapped trajectory, one might be tempted to construct a postprocessing scheme to remove this noise altogether. However, such sanitation is highly nontrivial and its advantage over performing simulations at constant volume is not evident. For this reason, we instead take a toroidal view of the PBCs and treat the wrapped particle dynamics as the “true” dynamics corresponding to the given simulation

ensemble, despite the presence of multiplicative noise. If the multiplicative noise associated with barostat position rescaling is considered an issue, we recommend constant-volume simulations instead of the complex and somewhat arbitrary postprocessing of trajectories generated at constant pressure.

Reassuringly, we expect published studies of diffusion coefficients to be minimally affected by the choice of unwrapping scheme. For one, early simulations tended to be short, making diffusion far beyond the central box and thus unwrapping artifacts rare, in particular for larger molecules. In addition, diffusion coefficients were usually estimated by least-squares fitting to the MSD, which we found to be quite robust at sufficiently long times (section 5.4).

Although current MD simulations of large biological molecules are only minimally affected by the shortcomings of the LAT scheme reported here, we expect our findings to become crucial for the proper analysis of future simulation trajectories on the time scales of milliseconds and beyond. For NPT simulations, we recommend the use of large boxes, for which the time to diffuse over multiple box dimensions is large and the position rescaling effects are small. The latter follows from the decay in the relative fluctuations of the box volume  $V$  with system size, namely  $\beta\langle(V - \langle V \rangle)^2\rangle = \chi_T\langle V \rangle$ , where  $\chi_T$  denotes the isothermal compressibility.<sup>32</sup>

For precision calculations of diffusion coefficients and related quantities, one may want to resort to NVT simulations. The choice of an NVT ensemble is advisable in particular for long simulations with small boxes of highly compressible systems with low viscosity, where the relative box-size fluctuations are large and particles can diffuse over many box widths. At constant volume, the lattice and toroidal view of PBCs coincide and unwrapping is unambiguous. If needed, the results from NVT simulations at different volumes can be interpolated to the targeted pressure or rigorously combined into weighted samples of an NPT ensemble.<sup>33,34</sup> An added advantage of working at constant volume is that the box size and shape entering the large finite-size corrections of translational diffusion coefficients<sup>12,28</sup> are well defined, whereas for NPT conditions one usually resorts to averages.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.3c00308>.

Supporting Movie S1 compares the time evolution of 2D trajectories generated via the TOR and LAT schemes (MP4)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Jakob Tómas Bullerjahn** – Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; [orcid.org/0000-0002-2459-219X](https://orcid.org/0000-0002-2459-219X); Email: [jakob.bullerjahn@biophys.mpg.de](mailto:jakob.bullerjahn@biophys.mpg.de)

**Gerhard Hummer** – Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany; [orcid.org/0000-0001-7768-746X](https://orcid.org/0000-0001-7768-746X); Email: [gerhard.hummer@biophys.mpg.de](mailto:gerhard.hummer@biophys.mpg.de)

## Authors

**Sören von Bülow** – Structural Biology and NMR Laboratory, Linderström-Lang Centre for Protein Science, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark; [orcid.org/0000-0001-7125-5973](https://orcid.org/0000-0001-7125-5973)

**Maziar Heidari** – Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany; [orcid.org/0000-0002-8081-6602](https://orcid.org/0000-0002-8081-6602)

**Jérôme Hénin** – Laboratoire de Biochimie Théorique UPR 9080, Institut de Biologie Physico-Chimique, CNRS and Université Paris-Cité, 75005 Paris, France; [orcid.org/0000-0003-2540-4098](https://orcid.org/0000-0003-2540-4098)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jctc.3c00308>

## Funding

Open access funded by Max Planck Society.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

We thank Dr. Balázs Fábíán for insightful discussions. J.T.B., M.H., and G.H. acknowledge financial support by the Max Planck Society. J.H. acknowledges support from the French Agence Nationale de la Recherche under grant DYNAMO (ANR-11-LABX-0011) and from Laboratoire International Associé UIUC-CNRS. S.v.B. acknowledges support from the EMBO Postdoctoral Fellowship ALTF 810-2022.

## ■ REFERENCES

- Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **1980**, *72*, 2384–2393.
- von Bülow, S.; Bullerjahn, J. T.; Hummer, G. Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure. *J. Chem. Phys.* **2020**, *153*, 021101.
- Kulke, M.; Vermaas, J. V. Reversible unwrapping algorithm for constant-pressure molecular dynamics simulations. *J. Chem. Theory Comput.* **2022**, *18*, 6161–6171.
- Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
- Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M.; Kollman, P. A. *AMBER 2018*; University of San Francisco, 2018.
- Zwanzig, R. Time-correlation functions and transport coefficients in statistical mechanics. *Annu. Rev. Phys. Chem.* **1965**, *16*, 67–102.
- Hénin, J. *qwrap – Fast PBC wrapping and unwrapping for VMD*; GitHub, 2016. <https://github.com/jhenin/qwrap> (accessed 2023-05-08).
- Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; in 't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; Shan, R.; Stevens, M. J.; Tranchida, J.; Trott, C.; Plimpton, S. J. LAMMPS – a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.



- (9) Fiorin, F.; Klein, M. L.; Héning, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362.
- (10) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Héning, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130.
- (11) Roy, R.; Olver, F. W. J. Elementary functions. In *NIST Handbook of Mathematical Functions*; Olver, F. W. J., Lozier, D. W., Boisvert, R. F., Clark, C. W., Eds.; Cambridge University Press: New York, NY, 2010; pp 103–134.
- (12) Yeh, I.-C.; Hummer, G. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions. *J. Phys. Chem. B* **2004**, *108*, 15873–15879.
- (13) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.
- (14) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (15) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.
- (16) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (17) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (18) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.
- (19) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.
- (20) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327–341.
- (21) Hockney, R. W.; Eastwood, J. W. *Computer Simulation Using Particles*; CRC Press: New York, NY, 1988.
- (22) Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **1984**, *81*, 511–519.
- (23) Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (24) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (25) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (26) Bullerjahn, J. T.; Hummer, G. Maximum likelihood estimates of diffusion coefficients from single-particle tracking experiments. *J. Chem. Phys.* **2021**, *154*, 234105.
- (27) Bullerjahn, J. T.; von Bülow, S.; Hummer, G. Optimal estimates of self-diffusion coefficients from molecular dynamics simulations. *J. Chem. Phys.* **2020**, *153*, 024116.
- (28) Vögele, M.; Hummer, G. Divergent diffusion coefficients in simulations of fluids and lipid membranes. *J. Phys. Chem. B* **2016**, *120*, 8722–8732.
- (29) Giorgino, T.; Héning, J.; Hoermann, J.; Kulke, M.; Lenz, O.; Mura, C.; Rogers, D. M.; Saam, J. *PBCTools plugin for VMD, version 3.1*; GitHub, 2022. <https://github.com/frobnitzem/pbctools> (accessed 2023-05-08).
- (30) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
- (31) Vestergaard, C. L.; Blainey, P. C.; Flyvbjerg, H. Optimal estimation of diffusion coefficients from single-particle trajectories. *Phys. Rev. E* **2014**, *89*, 022726.
- (32) Hill, T. I. *An Introduction to Statistical Thermodynamics*; Dover Publications: New York, NY, 1986.
- (33) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987.
- (34) Wood, W. W. Monte Carlo calculations for hard disks in the isothermal-isobaric ensemble. *J. Chem. Phys.* **1968**, *48*, 415–434.

## Recommended by ACS

### When MELD Meets GaMD: Accelerating Biomolecular Landscape Exploration

Marcelo Caparotta and Alberto Perez

DECEMBER 01, 2023

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

### Robust Estimation of Position-Dependent Anisotropic Diffusivity Tensors from Stochastic Trajectories

Tiago S. Domingues, Amir Haji-Akbari, et al.

JUNE 01, 2023

THE JOURNAL OF PHYSICAL CHEMISTRY B

READ 

### Coarse-Grained Modeling Using Neural Networks Trained on Structural Data

Mikhail Ivanov, Alexander P. Lyubartsev, et al.

SEPTEMBER 15, 2023

JOURNAL OF CHEMICAL THEORY AND COMPUTATION

READ 

### Soft Matter under Pressure: Pushing Particle-Field Molecular Dynamics to the Isobaric Ensemble

Samiran Sen, Michele Cascella, et al.

MARCH 28, 2023

JOURNAL OF CHEMICAL INFORMATION AND MODELING

READ 

Get More Suggestions >