THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

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

*Published as part of The Journal of Physical Chemistry virtual special issue "Early-Career and Emerging Researchers in Physical Chemistry Volume 2".*

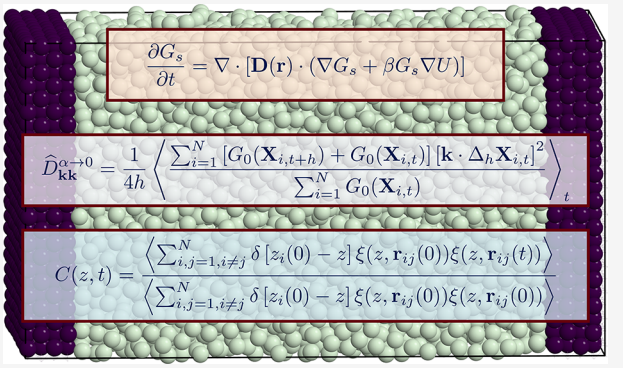
Tiago S. Domingues, Ronald R. Coifman, and Amir Haji-Akbari*

Cite This: *J. Phys. Chem. B* 2023, 127, 8644–8659

Read Online

ACCESS | Metrics & More | Article Recommendations

**ABSTRACT:** Confinement breaks translational and rotational symmetry in materials and makes all physical properties functions of position. Such spatial variations are key to modulating material properties at the nanoscale, and characterizing them accurately is therefore an intense area of research in the molecular simulations community. This is relatively easy to accomplish for basic mechanical observables. Determining spatial profiles of transport properties, such as diffusivity, is, however, much more challenging, as it requires calculating position-dependent autocorrelations of mechanical observables. In our previous paper (Domingues, T.S.; Coifman, R.; Haji-Akbari, A. *J. Phys. Chem. B* **2023**, *127*, 5273 10.1021/acs.jpcb.3c00670), we analytically derive and numerically validate a set of filtered covariance estimators (FCEs) for quantifying spatial variations of the diffusivity tensor from stochastic trajectories. In this work, we adapt these estimators to extract diffusivity profiles from MD trajectories and validate them by applying them to a Lennard-Jones fluid within a slit pore. We find our MD-adapted estimator to exhibit the same qualitative features as its stochastic counterpart, as it accurately estimates the lateral diffusivity across the pore while systematically underestimating the normal diffusivity close to hard boundaries. We introduce a conceptually simple and numerically efficient correction scheme based on simulated annealing and diffusion maps to resolve the latter artifact and obtain normal diffusivity profiles that are consistent with the self-part of the van Hove correlation functions. Our findings demonstrate the potential of this MD-adapted estimator in accurately characterizing spatial variations of diffusivity in confined materials.




## I. INTRODUCTION

Confinement is ubiquitous in nature and can change the physics and chemistry of matter in meaningful ways. For instance, it can shift phase boundaries[1−5] and glass-transition temperatures,[2,6,7] alter mechanical,[8,9] dielectric,[10,11] and transport[12−16] properties of materials, and result in the formation of new phases.[17,18] Moreover, confinement is commonly employed to modulate the kinetics and mechanism of rare events, such as nucleation[19,20] and cavitation.[21,22] The effect of confinement is not limited to atomic and molecular scales and can impact the mesoscale behavior of soft matter[23] and biological[24] systems. Beyond the potential quantum effects responsible for some emergent phenomena at the nanoscale, confinement generally affects materials by breaking translational symmetry, thereby making all physical properties functions of position. Characterizing the confinement-induced dependence of material properties on position is crucial to understanding material behavior and function under confinement and is therefore of interest from both a theoretical and a practical perspective.

In recent decades, molecular simulations[25] have proven indispensable in characterizing the role of confinement in modulating material properties.[26−33] One of the main advantages of molecular simulations is their ability to estimate spatial profiles of physical properties at resolutions that are not achievable through experiments. However, obtaining these profiles is straightforward for only static properties, which are mechanical observables that can be unambiguously estimated for each configuration. Spatial profiles of such properties can

Received: May 27, 2023
Revised: August 11, 2023
Published: September 27, 2023







https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

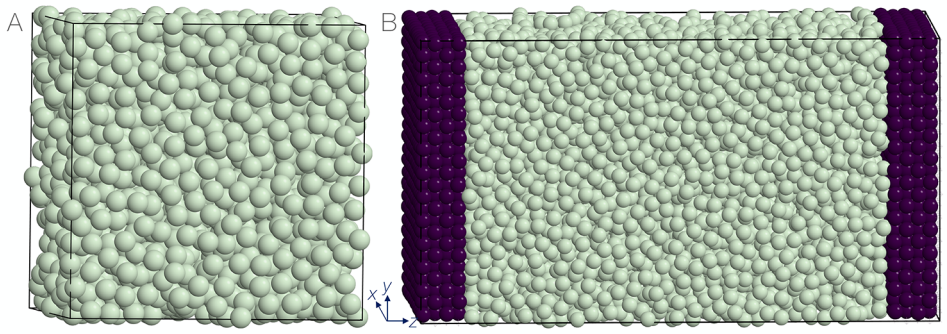



**Figure 1.** Schematic representation of the simulation box in (A) bulk and (B) confined geometries prepared using `injavis`.[57] A and B particles are depicted in light green and dark purple, respectively.

then be estimated by partitioning the simulation box into sufficiently small bins and estimating time averages of the relevant quantities within each bin. In contrast, dynamic properties such as transport coefficients can be estimated only from autocorrelations of mechanical observables using either the Green−Kubo[34] or the Helfand framework.[35] Neither of these frameworks has been rigorously generalized to systems with broken translation symmetry. Even assuming their validity at a local level, there is ambiguity in rigorously accounting for the contributions of molecules or particles that move across different bins during a specific observation window. Therefore, even though several *ad hoc* extensions[36−39] of bulk-based approaches have been proposed, they have by and large not been proven analytically or validated numerically.

A notable example of this challenge is diffusivity, which, in bulk systems, can be readily estimated from either the integral of the velocity autocorrelation function (VACF)[40] or the asymptotic first derivative of the mean-squared displacement (MSD).[41] Both of these approaches directly follow from the analytical solution of the Fokker−Planck equation,[42,43] which is satisfied by $G_s(\mathbf{r}, t)$, the self-part of the van Hove correlation function, in bulk systems for sufficiently long $t$ values. Under confinement, however, $G_s(\mathbf{r}, t|\mathbf{r}_0, 0)$ satisfies the Smoluchowski equation, which is a generalization of the Fokker−Planck equation with position-dependent diffusivity and the potential of mean force (PMF). Since the Smoluchowski equation lacks a general analytical solution, estimating local diffusivity from single-particle trajectories is not trivial. Moreover, the $G_s(\mathbf{r}, t|\mathbf{r}_0, 0)$ values computed from molecular dynamics (MD) trajectories satisfy only the Smoluchowksi equation when $t$ is sufficiently large. Both of these factors make it difficult to devise simple rigorous estimators that rely on the local validity of the Fokker−Planck equation within each bin. Despite several attempts to construct more rigorous estimators using sophisticated approaches such as Markov state models and Bayesian inference,[31,44−52] most authors have resorted to utilizing different flavors of *ad hoc* MSDs and VACFs.[32,33,36−39] While the profiles computed from such approaches provide semiquantitative descriptions of dynamics in confined materials, they are seldom validated to ensure their consistency with the $G_s(\mathbf{r}, t|\mathbf{r}_0)$ values computed from MD trajectories.

In an earlier paper,[53] we analytically derived and numerically validated a collection of filtered covariance estimators (FCEs) for estimating diffusivity from stochastic trajectories whose displacement statistics are consistent with the Smoluchowski equation. These estimators are related to what is known in the statistics literature as Nadaraya-Watson-type estimators,[54,55] which are widely used for estimating parameters of stochastic differential equations[56] due to their ease of implementation and versatility. In this article, we build upon our previous work by successfully adapting the FCE estimators developed therein to single-particle trajectories obtained from MD simulations. We validate our method by demonstrating its ability to provide accurate estimates of diffusivity in the bulk (consistent with what one would obtain from MSD), and as a test case, we apply it to quantify spatial variations in diffusivity in a Lennard-Jones (LJ) liquid within a slit pore. By regenerating stochastic differential equation (SDE) trajectories based on the computed PMF and diffusivity profiles and comparing the $G_s(\mathbf{r}, t|\mathbf{r}_0)$ values obtained from such trajectories with those obtained from MD, we confirm that our estimator performs reasonably well across the board. We observe only minor deviations in the normal mobility of the particles in the vicinity of the wall. Such discrepancies arise due to a fundamental limitation of covariance-based estimators that makes them susceptible to temporal discretization errors near impermeable boundaries, as identified in our earlier work. We propose a simulated annealing approach based on diffusion maps to mitigate such discrepancies and obtain more accurate normal diffusivity profiles without a need to shorten our observation window.

This paper is organized as follows. Section II.A outlines the details of the system setup and molecular dynamics simulations. The proposed procedure for adapting the FCE estimators of ref 53 to MD trajectories is discussed in Section II.B. Section II.C provides technical details of the SDE integration for validating the computed diffusivity profiles. Section III presents our results, while Section IV is reserved for concluding remarks.

## II. METHODS

**II.A. System Description and Molecular Dynamics Simulations.** We assess the performance of the adopted FCE estimators by simulating a single-component standard Lennard-Jones (LJ) liquid in two different geometries, namely, in the bulk (Figure 1A) and within a slit pore (Figure 1B). Bulk simulations are conducted solely for the purpose of ensuring consistency since the diffusivity of the bulk liquid can be readily estimated via alternative means, such as MSD. The confined geometry, however, serves as a simple but nontrivial test case for benchmarking and validating the estimator and is comprised of an LJ liquid sandwiched between flexible walls whose constituent atoms are kept together via harmonic bonds. Liquid particles, A, interact with wall particles, B, via the LJ potential with $\epsilon_{AB} = 0.3_{AA}$ and $\sigma_{AB} = \sigma_{AA}$. All LJ interactions are truncated at a distance of $r_c = 2.5\sigma_{AA}$. Both liquid and wall


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

particles have identical masses, $m_A = m_B = m$. MD simulations are conducted using LAMMPS at a reduced temperature of $T^* = kT/\epsilon_{AA} = 1$. Equations of motion are integrated using the velocity Verlet algorithm with a time step of $\Delta t^* = \Delta t/\sigma_{AA}\sqrt{\epsilon_{AA}/m_A} = 2.5 \times 10^{-3}$. All simulations are carried out in the canonical ($NVT$) ensemble with temperature controlled using the Nosé−Hoover thermostat[58,59] with a time constant of 0.25 in reduced LJ units.

In both geometries, the system is initialized using the following procedure. First, particles of type A are placed on a face-centered cubic (FCC) lattice of the appropriate number density. For the slit pore geometry, particles of type B are also placed on the same lattice but within a different part of the simulation box. Harmonic bonds of length $r_0 = 1.08\sigma_{AA}$ and spring constant $k_s = 500\epsilon_{AA}/\sigma_{AA}^2$ are added between the $B$ particles, the distance of which falls within the $[1.07\sigma_{AA}, 1.09\sigma_{AA}]$ interval. The arising configuration is then used as the starting point of a long MD trajectory carried out for $5 \times 10^6$ time steps. In both bulk and slit pore geometries, the FCC lattice melts within a few thousand time steps. We save the liquid configuration every $5 \times 10^5$ steps, randomize the velocities of all particles in accordance with the Boltzmann distribution, and conduct an equilibration simulation for $10^6$ steps. We then launch production runs from end points of such trajectories for $5 \times 10^6$ steps, with snapshots recorded every 50 steps for postprocessing.

**II.B. Estimation of Spatial Profiles of Diffusivity.** As mentioned in Section I, the procedure for obtaining spatial profiles of static properties from MD trajectories is rigorous and well established. First, an equilibrium number density profile $\rho(\mathbf{r})$ is obtained as

$$\rho(\mathbf{r}) = \left\langle \sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \right\rangle \tag{1}$$

where $N$ is the total number of particles in the system. Note that $\mathcal{F}(\mathbf{r}) = -\beta^{-1} \ln \rho(\mathbf{r})$ is the associated potential of mean force (PMF) and $p_0(\mathbf{r}) = \rho(\mathbf{r})/\int \rho(\mathbf{s})\, d^3\mathbf{s}$ is the equilibrium probability density. The intensity profile of a mechanical observable $\mathcal{A}$ that is well defined on a per-particle level (such as potential energy) can then be estimated as

$$\mathcal{A}(\mathbf{r}) = \frac{1}{\rho(\mathbf{r})} \left\langle \sum_{i=1}^{N} \mathcal{A}_i \delta(\mathbf{r}_i - \mathbf{r}) \right\rangle \tag{2}$$

In practice, eqs 1 and 2 are estimated by partitioning the simulation box into smaller bins and estimating a per-bin average of the desired quantity. Mathematically, this implies an approximation of the delta function with the indicator function of each bin.

When it comes to transport coefficients, such as diffusivity, the procedure is more complex. In bulk systems and for observation windows that are sufficiently long, the behavior of single-particle trajectories is governed by the Fokker−Planck equation[42,43]

$$\frac{\partial G_s(\mathbf{r}, t|\mathbf{r}_0, 0)}{\partial t} = \nabla \cdot [\mathbf{D} \cdot \nabla G_s(\mathbf{r}, t|\mathbf{r}_0, 0)] \tag{3}$$

which has the following analytical solution:

$$G_s(\mathbf{r}, t|\mathbf{r}_0, t) = \frac{1}{\sqrt{\det(4\pi \mathbf{D} t)}} e^{-(\mathbf{r}-\mathbf{r}_0)^T \mathbf{D}^{-1} (\mathbf{r}-\mathbf{r}_0)/4t} \tag{4}$$

Here, $\mathbf{D}$ is the self-diffusive tensor, which is isotropic in most fluids. Using this analytical solution, $\mathbf{D}$ can be estimated from either the integral of the velocity autocorrelation function[40]

$$\mathbf{D} = \int_0^{+\infty} \langle \mathbf{v}(t)\, \mathbf{v}^T(0) \rangle\, dt \tag{5}$$

or the first derivative of the mean-square displacement:

$$\mathbf{D} = \lim_{t \to \infty} \frac{\langle [\mathbf{r}(t) - \mathbf{r}(0)][\mathbf{r}(t) - \mathbf{r}(0)]^T \rangle}{2t} \tag{6}$$

In confined systems, single-particle trajectories obey the Smoluchowski equation, which accounts for the effect of position-dependent diffusivity $\mathbf{D}(r)$ and PMF $U(\mathbf{r})$:

$$\frac{\partial G_s}{\partial t} = \nabla \cdot [\mathbf{D}(\mathbf{r}) \cdot (\nabla G_s + \beta G_s \nabla U)] \tag{7}$$

Unlike eq 3, eq 7 does not have an analytical solution. Consequently, neither of the frameworks given by eqs 5 and 6 can be readily generalized to estimate position-dependent diffusivity profiles in translationally anisotropic systems. An alternative approach, which is conceptually feasible, is to use a Bayesian scheme to infer the diffusivity profile from the $G_s(\mathbf{r}, t|\mathbf{r}_0)$ values obtained from MD trajectories. One approach that stands out is the use of Markov state models. This approach involves partitioning the simulation box into smaller bins and discretizing $G_s(\mathbf{r}, t|\mathbf{r}_0, 0)$ to obtain a transition matrix for jumps between the bins.[44,52] The Smoluchowski operator is then temporally discretized to relate the transition matrix to local diffusivity. Despite being mathematically rigorous, such approaches are tedious to implement, as estimating $G_s(\mathbf{r}, t|\mathbf{r}_0, 0)$ generally requires spatial binning in a space with a maximum of six dimensions, and sufficient statistics might not exist for sampling different regions of such a space. Moreover, due to large spatial variations in particle mobility across the system, the statistical accuracy of $G_s(\mathbf{r}, t|\mathbf{r}_0, 0)$ might be uneven for different $\mathbf{r}$ values, and this issue is not easy to mitigate due to the constraint that a single propagation time scale needs to be utilized for the entire system.

Due to these challenges, an overwhelming majority of authors have employed heuristic estimators based on *ad hoc* generalizations of the notion of a mean-squared displacement to spatial bins and estimated the local lateral diffusivity from the asymptotic slope of the per-bin MSD. There are several issues with such approaches. First and foremost, the convergence of such estimators to the true local diffusivity has yet to be established analytically or validated numerically. In fact, it has been previously demonstrated that such generalized MSDs do not exhibit the generic behavior of bulk MSD even over long time scales due to interlayer mixing.[32] Second, there is ambiguity in how to define the mean-squared displacement for an open system, e.g., spatial bins that exchange particles with one another. Finally, such approaches can at best estimate the lateral components of the diffusivity tensor (e.g., the $xx$ and $yy$ components in the case of one-dimensional confinement along the $z$ direction).

In our earlier paper,[53] we constructed a series of filtered covariance estimators (FCEs) that are closely related to Nadaraya-Watson-type estimators in statistics and that can accurately estimate anisotropic and position-dependent diffusivity from stochastic trajectories whose displacement statistics follow the Smoluchowski equation. It must be noted


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

that individual realizations of such trajectories can be generated using the following SDE,[45]

$$d\mathbf{X}_t = [\mathbf{D}(\mathbf{X}_t)\cdot\nabla \log \rho(\mathbf{X}_t) + \nabla\cdot\mathbf{D}(\mathbf{X}_t)]dt + \sqrt{2\mathbf{D}(\mathbf{X}_t)}\, d\mathbf{W}_t \tag{8}$$

Displacements along such trajectories over short observation windows are then processed using a filter function $\gamma$: $\mathbb{R}^3 \to \mathbb{C}$, a piecewise $C^2$ function that filters displacements along certain directions and within certain domains. According to Itô's Lemma, the temporal evolution of $\gamma(\mathbf{X}_t)$ can be expressed as[60]

$$\lim_{h\to 0} \frac{\langle|\gamma(\mathbf{X}_{t+h}) - \gamma(\mathbf{X}_t)|^2\rangle_{\mathbf{X}_t\approx p_0(\cdot)}}{2h} = \int \nabla\gamma^\dagger \mathbf{D}(\mathbf{r})\nabla\gamma p_0(\mathbf{r})\, d^3\mathbf{r} \tag{9}$$

In order to retrieve pointwise diffusivities from this general formalism, we employed filter functions of the form $\gamma_\mathbf{k}(\mathbf{r}) := f_\mathbf{k}(\mathbf{r})\, g(\mathbf{r}-\mathbf{r}_0)$ wherein $f_\mathbf{k}(\mathbf{r}) = e^{-i\alpha\mathbf{k}\cdot\mathbf{r}}$ (with $\mathbf{k}\in S^2$ being a unit vector) captures the anisotropy of the diffusivity tensor while $g(\mathbf{r})$ is a real-valued localization function centered at the origin. Upon some mathematical rearrangements, it can be demonstrated that

$$\int \mathbf{k}^T\mathbf{D}(\mathbf{r})\mathbf{k} p_0(\mathbf{r}) g(\mathbf{r}-\mathbf{r}_0)\, d^3\mathbf{r} = \lim_{h\to 0} \frac{\langle[\gamma_\mathbf{k}(\mathbf{X}_{t+h}) - \gamma_\mathbf{k}(\mathbf{X}_t)]^*[f_\mathbf{k}(\mathbf{X}_{t+h}) - f_\mathbf{k}(\mathbf{X}_t)]\rangle^{\Re}_{p_0}}{2\alpha^2 h} \tag{10}$$

Note that the left side of eq 10 will converge to $D_{\mathbf{kk}}(\mathbf{r}_0) = \mathbf{k}^T\mathbf{D}(\mathbf{r}_0)\mathbf{k}$ as $g(\mathbf{r}) \to \delta(\mathbf{r})$. By choosing a sufficiently short observation window $h > 0$ and a sufficiently narrow localization function, an FCE estimator for $D_{kk}(\mathbf{r}_0)$ can be constructed as

$$\hat{D}_{\mathbf{kk}} = \frac{1}{2\alpha^2 h}\left\langle \frac{\Re[\sum_{i=1}^N \Delta_h\gamma_\mathbf{k}^*(\mathbf{X}_{i,t})\, \Delta_h f_\mathbf{k}(\mathbf{X}_{i,t})]}{\sum_{i=1}^N g(\mathbf{X}_{i,t}-\mathbf{r}_0)}\right\rangle_t \tag{11}$$

where $N$ is the number of particles and $\Delta_h F(\mathbf{X}_t) := F(\mathbf{X}_{t+h}) - F(\mathbf{X}_t)$. The full diffusivity tensor can then be constructed by choosing a sufficient number of $m \geq 6$ suitable unit vectors and combining the respective estimates given by eq 11 through a least-squares approach. It can be demonstrated that the limiting behaviors of eq 11 at $\alpha \to 0$ will be given by

$$\hat{D}_{\mathbf{kk}}^{\alpha\to 0} = \frac{1}{4h}\left\langle \frac{\sum_{i=1}^N [G_0(\mathbf{X}_{i,t+h}) + G_0(\mathbf{X}_{i,t})][\mathbf{k}\cdot\Delta_h\mathbf{X}_{i,t}]^2}{\sum_{i=1}^N G_0(\mathbf{X}_{i,t})}\right\rangle_t \tag{12}$$

where $G_0(\mathbf{r}) = g(\mathbf{r}-\mathbf{r}_0)$. Hereafter, for notational simplicity, we will drop $\alpha \to 0^+$ from our expressions since this will be the only version of the estimator that will be used. Instead, we may write $\tilde{D}_{\mathbf{kk}}^h$ to highlight the dependency of the estimator on the time scale $h$, particularly when different values of the observation time scale are simultaneously utilized. It is noteworthy that eq 12 provides a natural means of defining a generalized MSD for open systems. More specifically, by replacing $g(\cdot)$ with the indicator function of each spatial bin, a generalized MSD can be constructed by adopting the convention that the displacements of particles that are present within the bin either at the beginning or at the end of the observation windows will contribute to the MSD, but at 50% weight. The weight will be 100% for particles located within the bin both at the beginning and at the end of the observation window.

Despite the rigorous nature of the estimators proposed in ref 53, they cannot be directly applied to trajectories obtained from MD simulations, which behave only diffusively (i.e., in accordance with the Smoluchowski equation) at sufficiently long time scales. At shorter time scales, what is generally known as the "ballistic regime" takes hold in which MSD scales quadratically with time (i.e., $\langle|\mathbf{X}_t - \mathbf{X}_0|^2\rangle \propto t^2$). Therefore, the duration of the observation window $h$ needs to be larger than $t_{tr}$, the characteristic time scale for transition from the ballistic regime to the diffusive regime. In bulk systems, $t_{tr}$ can be readily estimated by examining the slope of $\log\langle|\mathbf{X}_t - \mathbf{X}_0|^2\rangle$ vs $\log t$ and identifying the time at which that slope approaches unity. Applying a similar procedure to confined systems is not feasible not only since there is ambiguity in the proper definition of MSD but also because the log−log slope is not guaranteed to converge to unity. Moreover, $t_{tr}$ can be strongly dependent on position.

A suitable proxy for $t_{tr}$ in dense liquids is the cage escape time (i.e., the time that it takes for a particle to escape the cage formed by its first coordination shell). In the bulk, the first coordination shell corresponds to all neighbors that are closer to the central particle than $r_{cage}$, the first valley of the radial distribution function (RDF). In confined systems, however, the first valley of RDF can also depend on position, and one needs to employ a localized definition of RDF. Here, we use the $z$-dependent lateral RDF defined as

$$g(|\mathbf{r}^{\parallel}|, z) = \frac{1}{\rho(z)}\sum_{i<j=1}^N \langle\delta(|\mathbf{r}_{ij}^{\parallel}| - |\mathbf{r}^{\parallel}|)\times\delta(z_i - z)\delta(z_j - z)\rangle \tag{13}$$

Here, $\rho(z) := \langle\sum_{i=1}^N \delta(z_i - z)\rangle$, $\mathbf{r}^{\parallel} \equiv (x, y, 0)$, the projection of the position vector onto the $xy$ plane, and $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$. $r_{cage}(z)$ is then given by the first valley of $g(r, z)$, which can then be used to define a cage escape autocorrelation function (CEAF) given by

$$C(z, t) = \frac{\left\langle\sum_{i,j=1,i\neq j}^N \delta[z_i(0) - z]\, \xi(z, \mathbf{r}_{ij}(0))\, \xi(z, \mathbf{r}_{ij}(t))\right\rangle}{\left\langle\sum_{i,j=1,i\neq j}^N \delta[z_i(0) - z]\, \xi(z, \mathbf{r}_{ij}(0))\, \xi(z, \mathbf{r}_{ij}(0))\right\rangle} \tag{14}$$

with $\xi(z, \mathbf{r}) = H[r_{cage}(z) - \|\mathbf{r}\|]$ and $H(\cdot)$ being the Heaviside function. Intuitively, $C(z, t)$ corresponds to the fraction of the neighboring particles that remains within the first coordination shell of a central particle located at $z$ after $t$ has elapsed. We compute $C(z, t)$ for central particles belonging to each spatial bin and fit it to the stretched exponential profile $C(z, t) = \exp[-[t/\tau(z)]^{\alpha(z)}]$. We then use $\tau(z)$ as a proxy for $t_{tr}(z)$ in our estimator.

There is, however, one additional issue in adopting the estimators given by eqs 11 and 12 to MD trajectories due to the observation that in dense liquids there is always an intermediate caging regime between the ballistic and diffusive regimes. During this caging regime, which can be extremely long for glassy and deeply supercooled liquids,[61] particle mobility is hampered as particles are trapped within the cages formed by their nearest neighbors. As a result, there is always a lag in the development of the diffusive behavior, and even in bulk systems, the intercept of the linear fit for $\langle\|\mathbf{r}(t) - \mathbf{r}_0\|^2\rangle$ vs $t$ is almost always negative. Such behavior cannot be directly captured using the eqs 11 and 12 estimators which assume


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

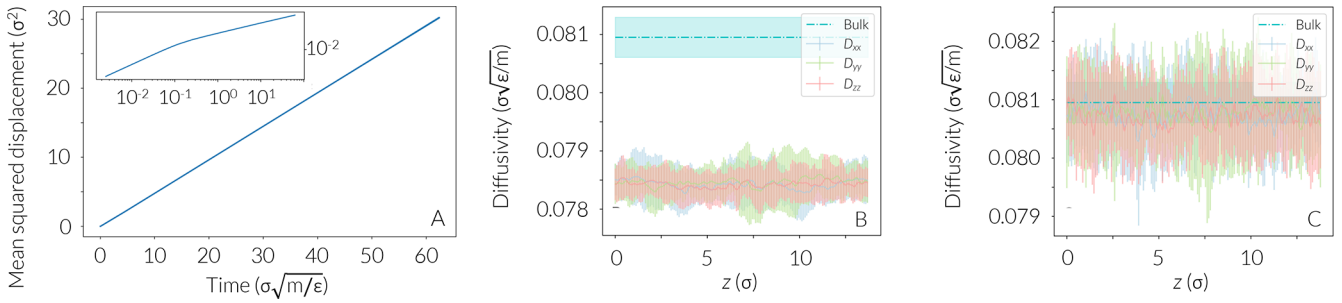



**Figure 2.** (A) Mean squared displacement for the bulk system, with the log−log plot shown in the inset. Local diffusivity estimates obtained from the estimators are given by (B) eq 12 and (C) eq 15. The 95% confidence interval for the bulk diffusivity obtained from MSD is shown in shaded cyan. Note that not accounting for the caging lag can result in a systematic underestimation of diffusivity.

diffusive behavior across all time scales. What is instead needed is to estimate local diffusivity from mobilities over two distinct observation windows $h_1 < h_2$ (both within the diffusive regime). In particular, we construct a composite estimator of the form

$$\tilde{D}_{\mathbf{kk}}^{h_1,h_2} = \frac{h_2 \hat{D}_{\mathbf{kk}}^{h_2} - h_1 \hat{D}_{\mathbf{kk}}^{h_1}}{h_2 - h_1} \tag{15}$$

which is mathematically identical to the original FCE estimator when applied to stochastic trajectories. When applied to MD trajectories, however, eq 15 ensures that only the relative mobility of particles is captured over observation windows during which particle displacements are already diffusive.

As for other features of the FCE estimator, we specifically employ the limiting estimator for $\alpha \to 0$ and given by eq 12. Similar to our earlier publication, we use a deterministic stencil of 13 vectors, which are the normalized versions of the vectors in the following set:

$$\mathbf{K} = \{(k_x, k_y, k_z): k_x, k_y \in \{0, \pm 1\}, k_z \in \{0, 1\}\}$$

We also choose $g(z) = (1/2\kappa)e^{-1}/(|1 - (z/2\kappa)^2|^2)$ as our localization function with $\kappa = 0.05\sigma_{AA}$. Estimates of all static and dynamic properties are obtained within bins that have a thickness of $0.0762\sigma_{AA}$.

**II.C. Details of SDE Integration.** In order to validate the statistical consistency of the computed profiles with displacement statistics obtained from MD, we use the open-source Julia package, `DifferentialEquations.jl`,[62] for numerically integrating the SDE given in eq 8 using the diffusivity profiles obtained from MD trajectories. The `DifferentialEquations.jl` package uses an explicit high-order Runge−Kutta discretization to integrate eq 8. In order to be consistent with our MD trajectories, we use a time step of $\delta t_s = 0.0025\sigma_{AA}/\sqrt{\epsilon_{AA}/m_A}$. From each spatial bin, $N_p\rho_k$ single-particle trajectories are initiated and conducted for $500t_{cage}(z)/\delta t_s$ steps, and the position of the particle is saved every 50 steps. Here, $N_p = 6912$ is the number of liquid particles, and $\rho_k$ is the average number density in bin $k$.

In order to use the numerical solver of the `DifferentialEquations.jl` package, we need to provide it with the functional forms of both $\mathbf{D}(z)$ and $\nabla \log \rho(z)$. We construct such a functional form of $\mathbf{D}(z)$ via a simple linear interpolation between successive pointwise estimates. A similar approach is employed for $\nabla \log \rho(z)$ with pointwise estimates obtained as

$$\nabla \log \rho(z) \approx \frac{\rho(z + \Delta z) - \rho(z - \Delta z)}{\rho(z)} \tag{16}$$

Here, $\rho(z)$ is obtained using eq 1. For SDE trajectories, it is also necessary to construct reliable extrapolations of $\mathbf{D}(z)$ and $\nabla\rho(z)$ for regions of the simulation box not visited in our MD simulations, as particles might still visit such regions, albeit with low probability. Therefore, we use a simple procedure when it comes to diffusivity, assuming that it remains constant at the last value that we estimate and then drops linearly to $\mathbf{0}$ as the particle hits the wall. It is, however, necessary to employ a more systematic approach for extrapolating the drift term due to the heightened sensitivity of particle trajectories to it. Motivated by the seminal work of Magda et al.,[29] we fit the last few estimates obtained from eq 16 to a power law, namely, $\log \rho(z) = A(z - z_w)^{-a}$, with $z_w$ being the position of the wall. Due to the large statistical uncertainties of $\rho(z)$ and $\nabla \log \rho(z)$ estimates obtained from eqs 1 and 16 in the immediate vicinity of the wall, we use a weighted fitting approach with the weight of each data point inversely proportional to its absolute variance.

## III. RESULTS AND DISCUSSION

**III.A. Application of the FCE Estimator.** We first assess the performance of our estimator in the bulk system, where diffusivity can be readily obtained from MSD. (We compute MSD using the method of sliding windows and estimate bulk diffusivity using the MSD values within the range of $37.5 \le t^* \le 62.5$.) As can be noted in Figure 2A, the transition from the ballistic regime to the diffusive regime occurs over a fairly short observation windows (i.e., at $t_{tr}^* \approx 1$). In order to ensure consistency, we partition the simulation box into 160 bins along the $z$ direction and compute $t_{cage}$ for each bin, as a proxy for $t_{tr}^* \approx 1$. As expected, the $t_{cage}$ values obtained for different bins are consistent with one another and are found to be within the same order of magnitude as the $t_{tr}$ inferred from Figure 2A. More specifically, we obtain $t_{cage}^* = 5.9534 \pm 0.1304$, and we use its mean as our observation window, $h$. Using the standard FCE estimator given by eq 12 results in a slight but systematic underestimation of diffusivity as shown in Figure 2B. This is expected considering the lag in transitioning into the diffusive regime as can be observed in Figure 2A. The diffusivity estimates obtained from eq 15 with $h_1 = h$ and $h_2 = 2h$ are, however, statistically indistinguishable from the true value (obtained from MSD). Regardless, both estimators (with or without lag) respect the rotational symmetry of the systems and predict $D_{xx}$, $D_{yy}$, and $D_{zz}$ values that are statistically indistinguishable from one another. These observations


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

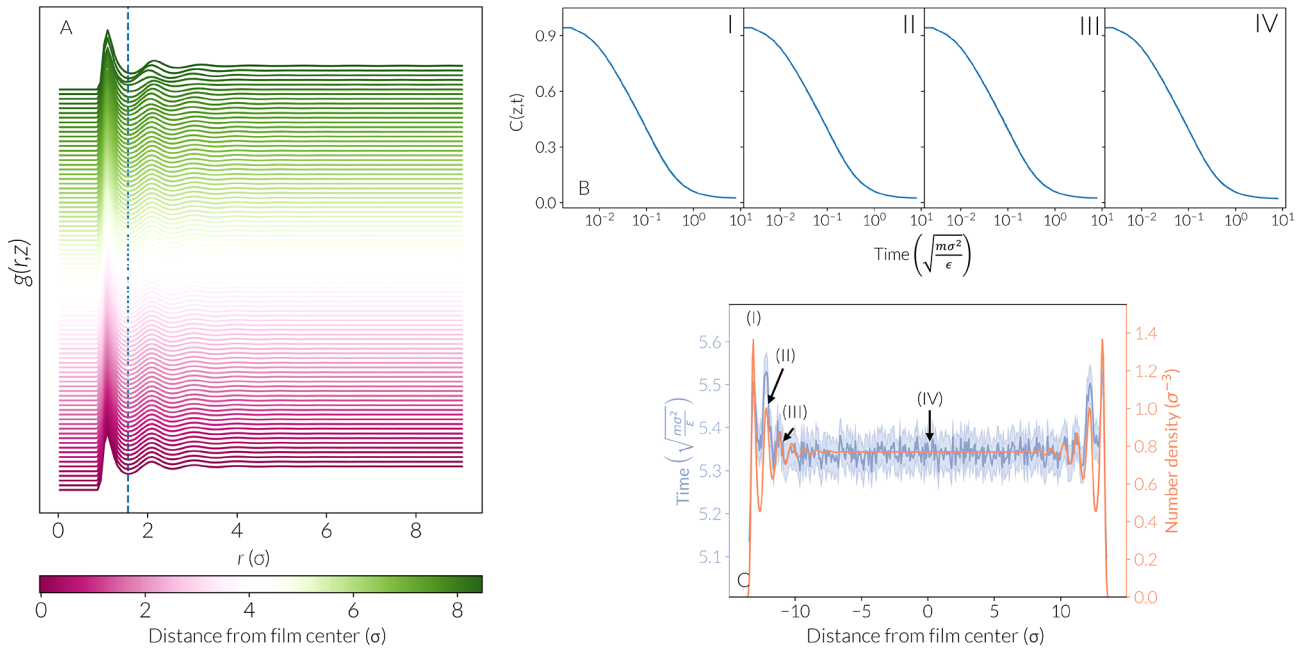



**Figure 3.** (A) Lateral radial distribution function as a function of $z$, the distance from the film center. Vertical dashed lines correspond to the approximate position of the first valley of RDF and are insensitive to $z$. (B) $C(z, t)$ for four representative spatial bins highlighted in (C) which depicts the number density and $t_{cage}$ as a function of $z$.

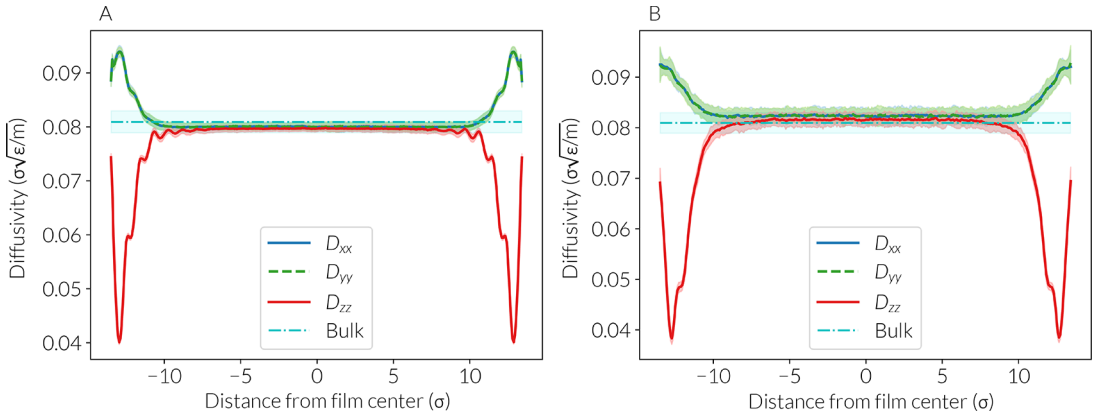



**Figure 4.** $\mathbf{D}(z)$ profiles estimated by using (A) eq 12 and (B) eq 15. The bulk diffusivity along with its 95% confidence interval is depicted in shaded cyan.

suggest that using an $h$ that is comparable in magnitude to $t_{tr}$ does not result in considerable discretization errors, and as such, FCE estimators can be readily applied to trajectories obtained from MD simulations.

We then apply the estimator to the more relevant and nontrivial case of a slit pore. As can be noted in Figure 3C, the existence of the confining walls makes the liquid structured, as evident from the multiple peaks in $\rho(z)$. Such structuring is not surprising and is a hallmark of liquids under confinement. Nonetheless, the locus of the first valley of RDF is not sensitive to such structuring, as can be seen in Figure 3A, and is almost constant at $r_{cage} \approx 1.5603\sigma_{AA}$. Indeed, the characteristic size of the first coordination shell within a liquid is governed by short-range steric and energetic interactions and is not expected to be very sensitive to temperature and pressure. We then use eq 14 to compute $C(z, t)$ for all spatial bins. Figure 3B depicts representative $C(z, t)$ values for the first three density peaks as well as the center of the film. We then compute $t_{cage}(z)$ by fitting each $C(z, t)$ to a stretched exponential. The computed $t_{cage}(z)$ profile is depicted in Figure 3C. It is noteworthy that spatial oscillations in the cage escape time follow those of the number density. Moreover, cage escape times are not very sensitive to $z$ and change by less than 10%.

By choosing an observation window of $h_1(z) := t_{cage}(z)$ and $h_2(z) = 2h_1(z)$ for each spatial bin, we use eq 15 to estimate $\mathbf{D}(z)$. The computed diffusivity profile is depicted in Figure 4B. Similar to the bulk, using eq 12 results in a slight underestimation of diffusivity as can be seen in Figure 4A. As expected, the diffusivity tensor is both isotropic and independent of $z$ in the center of the film as the estimated $D_{xx}(z)$, $D_{yy}(z)$, and $D_{zz}(z)$ profiles are statistically indistinguishable therein and fluctuate around an average value. In other words, the film center exhibits bulklike behavior due to its sufficiently large distance from the confining walls. In order to ensure the quantitative accuracy of the diffusivity estimates at the film center, however, we conduct $NVT$ simulations of the bulk liquid with number densities within the range of $[\bar{\rho}^* \pm \delta\rho^*]$, wherein $\bar{\rho}^* = 0.7691$ and $\delta\rho^* = 0.0015$ are the mean


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

and standard deviation of number density, respectively, for $|z| < 1.14\sigma_{AA}$. As can be seen in Figure 4, the diffusivity estimates obtained from the FCE estimator agree very well with those computed from MSD (highlighted in shaded cyan in Figure 4).

Upon approaching the confining walls, the diffusivity tensor starts becoming anisotropic. Nonetheless, $D_{xx}(z)$ and $D_{yy}(z)$ remain statistically indistinguishable from one another. This is consistent with our expectation that the diffusivity tensor will have only two independent components in this geometry (due to the equivalence of $x$ and $y$ directions). Nonetheless, $D_{zz}$ estimates differ considerably from those of $D_{xx} = D_{yy}$ due to the anisotropy induced by the wall.

**III.B. Numerical Validation of the FCE Estimator.** While the computed profile satisfies all of our basic physical expectations (isotropy at the center and equivalence of $D_{xx}$ and $D_{yy}$), its consistency with mobility statistics of individual particles still needs to be rigorously validated. In order to do so, we feed the computed diffusivity profile into a numerical solver of the SDE given by eq 8. As discussed in Section II.C, we estimate the drift term using eq 16 and the power-law extrapolation depicted in Figure 5. For such stochastic

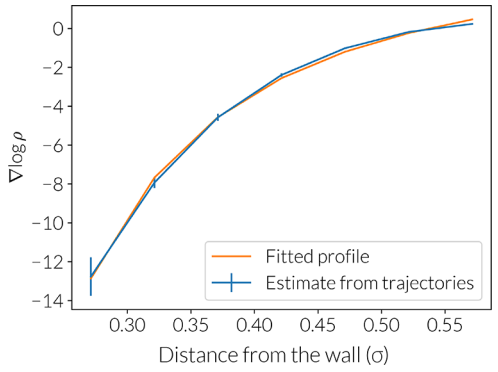



**Figure 5.** Power-law dependence of the drift term in the immediate vicinity of the wall.

trajectories, one can compute the self-part of the van Hove correlation function $G_s(r, z, t|0, z_0, 0)$ and compare it to those obtained from MD. In order to ensure better statistics, however, we compute the marginal distributions of $G_s(r, z, t|0, z_0, 0)$ given by

$$\tilde{G}_s(z_0, z, t) = \int_0^{+\infty} 2\pi r G_s(r, z, t|0, z_0, 0)\, dr \quad (17)$$

$$\hat{G}_s(z_0, r, t) = \int_{-\infty}^{+\infty} G_s(r, z, t|0, z_0, 0)\, dz \quad (18)$$

where $r$ refers to the lateral displacement of a molecule in the $xy$ plane from its initial position. Alternatively, the marginal probability densities given in eqs 17 and 18 can be referred to as normal and radial van Hove correlation functions, respectively. (Here, "normal" refers to mobility in the $z$ direction, i.e., perpendicular to the confining wall.) However, it must be noted that such a comparison is prone to small error due to the lag in transitioning from the ballistic regime to the diffusive regime. A more rigorous comparison can be made between a composite van Hove correlation function given by

$$G_s^{comp}(\mathbf{r}, t, t'|\mathbf{r}_0, 0) = \int G_s^{MD}(\mathbf{r}', t|\mathbf{r}_0, 0)\, G_s^{SDE}(\mathbf{r}, t'|\mathbf{r}', 0) d\mathbf{r}' \quad (19)$$

and the $G_s(\mathbf{r}, t + t'|\mathbf{r}_0, 0)$ obtained from MD. Note that $G_s^{comp}(\mathbf{r}, t, t'|\mathbf{r}_0, 0)$ is the convolution of the van Hove correlation functions obtained from MD and SDE for observation windows $t$ and $t'$, respectively. In other words, such an approach combines the actual displacement statistics over the time window $[0, t]$ with the expected displacement statistics due to diffusive motion over the time window $[t, t + t']$. Proper normal and radial averages of $G_s^{comp}(\cdot)$ can be formulated by using eqs 17 and 18, respectively.

As can be seen in Figure 6, the radial van Hove correlation functions obtained from MD and SDE for $t = h$ are already

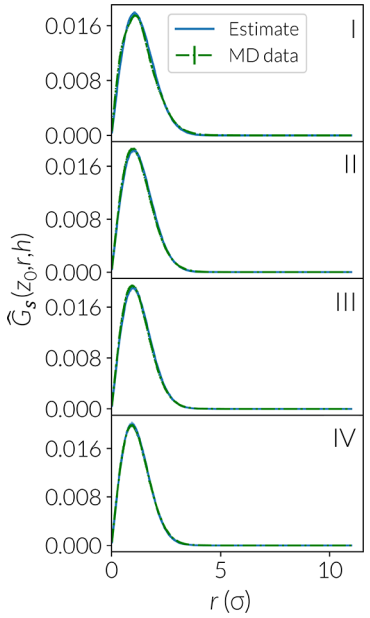



**Figure 6.** Comparison of the self-parts of the radial van Hove correlation functions computed from MD trajectories with those obtained from purely stochastic trajectories, consistent with the diffusivity estimate. Labels I–IV correspond to starting positions highlighted in Figure 3C.

statistically indistinguishable irrespective of the distance from the confining walls. This not only validates that our estimator accurately characterizes lateral diffusive behavior across the film but also implies that not using composite validation described by eq 19 does not lead to considerable errors. When it comes to the normal van Hove correlation functions, however, using the composite approach considerably improves the agreement between MD and SDE as can be seen in Figure 7. Nonetheless, even the composite $\tilde{G}_s$ values fail to agree fully with the ones obtained from MD except in the center of the film. More precisely, the estimated diffusivity profile correctly predicts the locations of peaks and valleys of $\tilde{G}_s(z, z_0, 2h)$ but not their amplitudes. Moreover, this discrepancy becomes increasingly worse upon approaching the wall.

In order to explain the observed disparities, we refer to our earlier findings in ref 53, wherein we demonstrated that covariance-based estimators, specifically the FCE estimator employed in this study, possess inherent limitations when it comes to accurately estimating the normal component of the diffusivity tensor in close proximity to hard boundaries when finite observation windows are employed. The root cause of this limitation can be intuitively attributed to the distortion of the Gaussian shape of the self-component of the van Hove correlation function upon encountering the hard boundary.


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

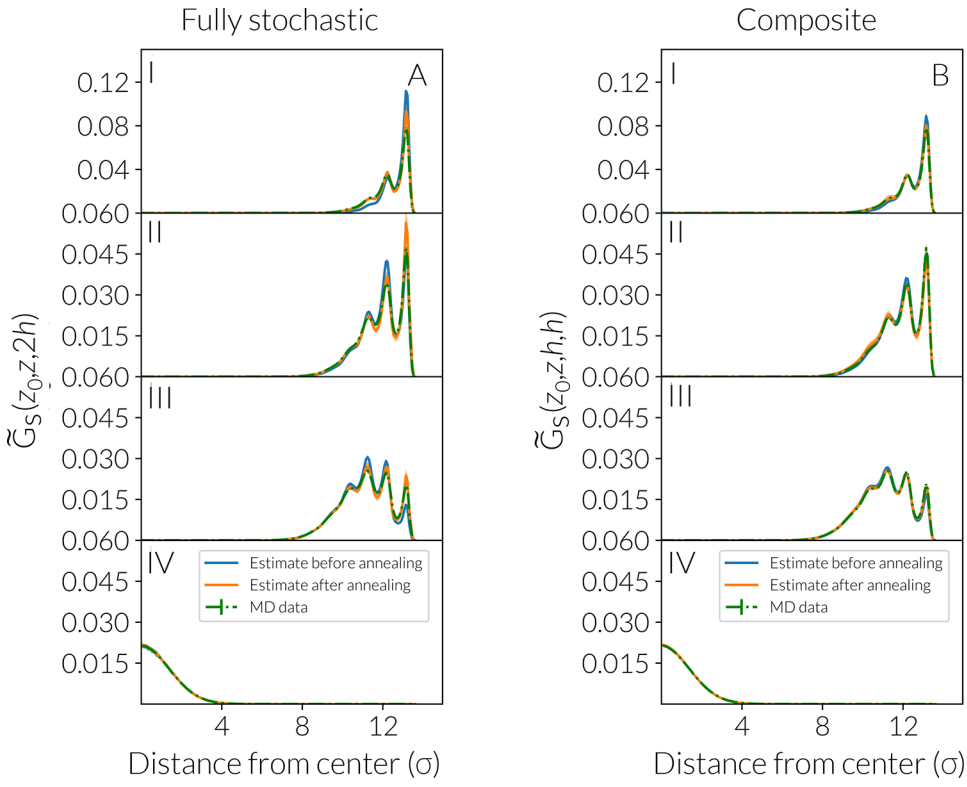



**Figure 7.** Comparison of the self-parts of the normal van Hove correlation functions computed from MD with those obtained (A) from purely stochastic trajectories, consistent with the diffusivity estimate and (B) via a composite approach described by eq 19 with $t = t' = h$. Labels I−IV correspond to starting positions highlighted in Figure 3C. The orange curves correspond to $\tilde{G}_s$ values obtained from stochastic trajectories integrated using the annealed diffusivity profile.

Consequently, deviations are expected in the estimated $D_{zz}(z)$ profile, whereas $D_{xx}(z)$ and $D_{yy}(z)$ remain unaffected. For systems exhibiting Smoluchowski-like behavior across all temporal scales, this issue can be mitigated by employing smaller observation windows. However, in the case of MD trajectories that solely display diffusive behavior over extended observation windows, such an adjustment is not feasible. Indeed, the observation window employed in this study represents the lower threshold beneath which single-particle trajectories will no longer exhibit diffusive characteristics. Thus, an alternative approach is imperative to rectify the estimates obtained from eq 15.

**III.C. Correction Scheme.** The correction scheme proposed here is based on defining a notion of distance between the $\tilde{G}_s(z, z_0, 2h)$ values obtained from MD and the composite $\tilde{G}_s^{comp}(z, z_0, h, h)$ constructed by combining MD and SDE data and using an efficient optimization scheme to minimize it. Let $d(P, Q)$ denote a smooth convex measure of distance between two probability density functions, $P$ and $Q$, satisfying the conditions $d(P, Q) = d(Q, P)$ and $d(P, Q) = 0 \Leftrightarrow P \equiv Q$. Formally, a typical correction scheme can be expressed as an optimization problem given by

$$D_{zz}^{*}(z) = \underset{D_{zz}(z)}{\operatorname{argmin}}\, d[\tilde{G}_s^{MD}(z_0, z, 2h), \tilde{G}_s^{comp}(z_0, z, h, h)] \tag{20}$$

In principle, one can solve eq 20 by adopting an iterative strategy in which $D_{zz}(z)$ is perturbed based on an optimization scheme and is then fed into an SDE integrator to compute the updated value of $d[\tilde{G}_s^{MD}(z_0, z, 2h), \tilde{G}_s^{comp}(z_0, z, h, h)]$. In practice, however, such an approach will be computationally expensive, as it will involve several rounds of computationally costly SDE integrations. We therefore adopt a more efficient alternative based on the concept of diffusion maps,[63] which provides an approximation of a Markov transition matrix for a particular diffusivity profile. Such a matrix quantifies the probability of transitioning between different spatial bins over a short observation window $\tau \ll h$ and is therefore a bin-discretized approximation of the self-part of the van Hove correlation function. The implementation details of the generalized algorithm for an arbitrary three-dimensional binning of the simulation domain are outlined in Algorithm 1. It is, however, necessary to note that for the specific problem at hand, binning is conducted in one dimension only, and as such, $\mathbf{D} = D_{zz}$ is a scalar.

Since diffusion maps are founded upon ideas similar to kernel-based estimators, their accuracy depends on the choice of the bandwidth parameter $\tau$, and it is only guaranteed in the limits of $\tau \to 0$ and $n_b \to \infty$. Therefore, even though the $\tau$ employed in Algorithm 1 needs to be considerably smaller than $h$, it needs to be chosen in conjunction with $n_b$ whereby the optimal $\tau$ is a decreasing function of $n_b$. Several sophisticated techniques have been proposed to identify suitable values for $\tau$. Nevertheless, in practical applications, the median value of the pairwise Mahalanobis distance matrix $\mathbf{A}$ is commonly considered an acceptable initial estimate for $\tau$. With such a choice, $\mathbf{P}^{D_{zz}\tau}$ can be computed from diffusion maps, and a reasonable approximation of $\tilde{G}_s^{st}(z_0, z, h)$ can be obtained as $\tilde{G}_s^{st}(z_0, z, h) = (\mathbf{P}^{D_{zz}\tau})^{N_s}$ wherein $N_s := h/\tau$. The principal advantage of this approach is its computational efficiency as contrasted with the numerical integration of an SDE. Moreover, unlike other applications of diffusion maps, such


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

**Algorithm 1** Target measure Mahalanobis diffusion maps.

1: **Inputs:**
   1. Positions of spatial bins $\{\mathbf{r}_i\}_{i=1}^{n_b}$.
   2. Pointwise estimates of number density and the diffusivity tensor at each bin, i.e., $\{\rho(\mathbf{r}_i)\}_{i=1}^{n_b}$ and $\{\mathbf{D}(\mathbf{r}_i)\}_{i=1}^{n_b}$, respectively.
   3. Time increment, $\tau$.
2: **Outputs:** a Markov matrix $\mathbf{P}^{\mathbf{D},\tau}$ with $\mathbf{P}_{ij}^{\mathbf{D},\tau}$ is the probability that a particle starting at bin $i$ at time $t$ reaches bin $j$ at $t+\tau$.
3: Compute $\mathbf{A}$, the matrix of pairwise Mahalanobis distances using:

$$\mathbf{A}_{ij} = \frac{1}{2}\sqrt{(\mathbf{r}_i - \mathbf{r}_j)^T \left[\mathbf{D}^{-1}(\mathbf{r}_i) + \mathbf{D}^{-1}(\mathbf{r}_j)\right](\mathbf{r}_i - \mathbf{r}_j)}$$

4: Compute the unnormalized kernel $\mathbf{K}^{(\tau)}$ and its associated row-summed vector $\mathbf{u}^{(\tau)}$ as:

$$\mathbf{K}_{ij}^{(\tau)} = \exp\left[-\frac{\mathbf{A}_{ij}^2}{\tau}\right]$$
$$u_i^{(\tau)} = \sum_{j=1}^{n_b} \mathbf{K}_{ij}^{(\tau)}$$

Note that by construction, $\mathbf{A}_{ij}^2$ has units of time.
5: Compute the normalization vector $\mathbf{n}_1$ as:

$$n_{1,i} = \frac{\sqrt{\rho(\mathbf{r}_i)}\,|\mathbf{D}(\mathbf{r}_i)|^{\frac{1}{4}}}{u_i^{(\tau)}}$$

Here $|\mathbf{D}|$ corresponds to the determinant of $\mathbf{D}$.
6: Use $\mathbf{n}_1$ to obtain the right-normalized kernel $\widehat{\mathbf{K}}^{(\tau)} = \mathbf{K}^{(\tau)}\,\mathrm{diag}(\mathbf{n}_1)$ where $\mathrm{diag}(\mathbf{n}_1)$ is a diagonal matrix whose diagonal elements are the components of $\mathbf{n}_1$.
7: Define the inverse row-summed vector $\mathbf{n}_2$ as:

$$n_{2,i} = \frac{1}{\sum_j \widehat{\mathbf{K}}_{ij}^{(\tau)}}$$

8: Obtain the output by left normalizing the kernel:

$$\mathbf{P}^{\mathbf{D},\tau} = \mathrm{diag}(\mathbf{n}_2)\widehat{\mathbf{K}}^{(\tau)}$$

9: **return** $\mathbf{P}^{\mathbf{D},\tau}$.

as dimensionality reduction,[64,65] there is no need to diagonalize the transition matrix. Rather, we are interested only in the fact that this construction provably approximates $e^{t\mathcal{L}_D^\dagger}$, where $\mathcal{L}_D^\dagger$ is the infinitesimal generator of the diffusion operator.[63,64] The computational efficiency of this construction therefore allows us to devise numerically efficient iterative methods to solve this problem (eq 20).

The next ingredient in our correction scheme is the selection of an appropriate distance measure, $d(\cdot,\cdot)$. Among the various options, the relative entropy emerges as a particularly suitable choice. Specifically, for two row-stochastic Markov matrices $\mathbf{P}$ and $\mathbf{Q}$, we define a vector comprising the row-wise distances as

$$H_i = d_{\mathrm{KL}}(P_i \| Q_i) = \sum_{j=1}^{n_b} P_{ij} \ln \frac{P_{ij}}{Q_{ij}} \tag{21}$$

We can then define a distance measure as

$$d(\mathbf{P}, \mathbf{Q}) = \frac{1}{n_b}\sum_{i=1}^{n_b} H_i \tag{22}$$

As discussed earlier, our binning is one-dimensional and is conducted only along the $z$ direction. More precisely, we partition the simulation domain into $n_b = 358$ evenly spaced bins with $\mathbf{P}_{ij} = \tilde{G}_s^{\mathrm{MD}}(z_j, z_i, 2h)$ and $\mathbf{Q}_{ij} = \tilde{G}_s^{\mathrm{comp}}(z_j, z_i, h, h)$. The stochastic contributions to $\tilde{G}_s^{\mathrm{comp}}$ are obtained via diffusion maps. It must also be noted that the correction scheme based on this particular choice of $d(\cdot,\cdot)$ can be nicely mapped onto a Bayesian framework, as discussed in Appendix A.

We then employ the simulated annealing technique, utilizing the Metropolis-Hastings Monte Carlo (MHMC) method, to solve the optimization problem specified by eq 20, wherein the objective function is treated as an energy that is gradually reduced by decreasing the temperature. Naturally, one would choose the distance measure of eq 22 as the energy function.


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

However, we empirically observe that doing so will result in artificial oscillations of $D_{zz}$ in the center of the film, possibly due to uncertainties in the computed $\tilde{G}_s^{MD}(z_0, z, t)$ values. Since our initial estimate of $D_{zz}$ within the film center is already precise (as validated from MSD), we augment $\mathcal{U}$ with a regularization term of the form

$$\mathcal{T} = \sum_{i=1}^{n_b-1} \overline{w}_i \|\mathbf{D}(\mathbf{r}_i) - \mathbf{D}(\mathbf{r}_{i+1})\|^2 \tag{23}$$

to penalize such oscillations. Here, $\|\cdot\|$ denotes the Frobenius norm of a matrix and reduces to $|D_{zz}(z_i) - D_{zz}(z_{i+1})|$ for the specific geometry considered in this work. Note that $\mathcal{T}(\mathbf{D})$ is akin to a "tension", similar to Tikhonoff regularization utilized for computing a pseudoinverse matrix.[66] One can then choose $\overline{w}_i$ values in such a manner that they vanish close to the walls (i.e., in regions where the bin-wise relative entropy is large. In this work, we set $\overline{w}_i = 0$ for the first 50 spatial bins next to each wall. The procedure employed for assigning $\overline{w}_i$ values of central bins is discussed in Algorithm 4. The energy function utilized for annealing is then modified to $\mathcal{U} + \lambda\mathcal{T}$, wherein $\lambda$ is a constant that tunes the relative strength of the tension term.

We start our simulated annealing calculations from the $D_{zz}(z)$ obtained from each MD trajectory using our FCE estimator, starting at an inverse temperature given by

$$\beta_a = \frac{A}{\max\limits_{1\le i\le n_b} H_i - \min\limits_{1\le i\le n_b} H_i} \tag{24}$$

For the findings presented in this article, we use a value of $A = 64$. Prior to generating trial moves, we first quantify the variability of diffusivity profiles obtained from independent MD trajectories. Specifically, let $\mathbf{D}_j(\mathbf{r}_i)$ denote the diffusivity estimate at $\mathbf{r}_i$ obtained from the $j$th MD trajectory. We construct $\mathbf{\Sigma} = \{\boldsymbol{\sigma}(\mathbf{r}_i)\}_{i=1}^{n_b}$, where $\sigma_{\alpha\beta}(\mathbf{r}_i)$ is the standard deviation of the $N_t$ independent estimates of $D_{\alpha\beta}(\mathbf{r}_i)$. Proper rescaling of these matrices provides the maximum step sizes of MC trial moves, which are generated according to the procedure described in Algorithm 2. In order to prevent the emergence of artificial oscillations, the proposed displacements to diffusivity are smoothed by convolving them with a normal distribution $\mathcal{N}(\mathbf{0}, 2{w_b}^2\mathbf{I})$, where $w_b$ denotes the characteristic width of a spatial bin and $\mathbf{I}$ is the identity matrix. In total, we conduct 10 annealing calculations, one per MD trajectory. All other implementation details, such as the quenching regime, are included in Algorithms 3, 4, and 5. We wish to stress that the implementation details outlined here are included for reproducibility purposes only and are not necessarily optimal. Therefore, it is possible to further optimize this procedure or to utilize alternative optimization strategies altogether.

Figure 8 depicts the outcome of a single annealing calculation initiated from a $D_{zz}(z)$ obtained from one of the 10 MD trajectories using our FCE estimator. As shown in Figure 8A, the objective function decreases sharply upon annealing, leading to a corrected diffusivity profile (Figure 8B) that differs significantly from the initial profile, particularly in the vicinity of the confining walls. Although the bin-wise relative entropies decrease considerably across the board, as depicted in Figure 8C, they do not descend to zero entirely. In particular, they deviate considerably from zero in the immediate vicinity of the walls partly due to inadequate statistics and considerable uncertainties as a result of extremely low number densities.

**Algorithm 2** Generating trial moves for simulated annealing

**Function** TrialMove $(\mathbf{\Sigma}, \mathbf{a})$.
**Inputs:** Standard deviation profile $\mathbf{\Sigma} = \{\boldsymbol{\sigma}(\mathbf{r}_i)\}_{i=1}^{n_b}$ and a rescaling vector $\mathbf{a} = (a_1, a_2, \cdots, a_{n_b})$ with non-negative entries.
**Outputs:** Proposed perturbation to the diffusivity profile, $\{\delta\mathbf{D}(\mathbf{r}_i)\}_{i=1}^{n_b}$.
Define vector $\mathbf{v}$ with components given by:

$$v_i = \frac{a_i}{\sum_{j=1}^{n_b} a_i}.$$

Note that $\mathbf{v} = \mathbf{a}/\|\mathbf{a}\|_0$.
Construct a collection of matrices $\overline{\boldsymbol{\sigma}}_i$ $(1 \le i \le n_b)$ given by:

$$\overline{\sigma}_{\alpha\beta,i} := Z\sigma_{\alpha\beta}(\mathbf{r}_i)$$

where $Z \sim \mathcal{N}(0,1)$ is a standard normal random variable.
For each bin, let $\delta\mathbf{D}(\mathbf{r}_i) := v_i \left[\overline{\boldsymbol{\sigma}}_i + \overline{\boldsymbol{\sigma}}_i^T\right]/2$.
**return** $\{\delta\mathbf{D}(\mathbf{r}_i)\}_{i=1}^{n_b}$.

**Algorithm 3** Quench$(\mathbf{\Delta}, \mathbf{\Sigma}, \delta S)$ - Temperature regime for simulated annealing

**Procedure** Quench$(\mathbf{\Delta}, \mathbf{\Sigma}, \delta S)$.
**Inputs:** An input diffusivity profile $\mathbf{\Delta} = \{\mathbf{D}_{in}(\mathbf{r}_i)\}_{i=1}^{n_b}$, a list of standard deviation matrices $\mathbf{\Sigma} = \{\boldsymbol{\sigma}(\mathbf{r}_i)\}_{i=1}^{n_b}$ and an initial step size $\delta S$.
**Outputs:** An updated diffusivity profile $\mathbf{\Delta}_{out} = \{\mathbf{D}_{out}(\mathbf{r}_i)\}_{i=1}^{n_b}$ and the final binwise relative entropy profile $\mathbf{H} = \{H_i\}_{i=1}^{n_b}$.
Use (24) to set the initial inverse temperature, $\beta$.
$c =$ **True**.
$n := 0$.
$\mathbf{\Delta}_0 := \mathbf{\Delta}$.
$T_{check} := 100$.
**while** $c$ **do**
    $[\mathbf{\Delta}_{n+1}, \mathbf{H}_{n+1}, a, b] =$MHMC$(\beta, \mathbf{\Delta}_n, \mathbf{\Sigma}, \delta S)$.
    $n := n + 1$.
    $\mathcal{E}_n := \mathcal{U}(\mathbf{\Delta}_n) + \lambda\mathcal{T}(\mathbf{\Delta}_n)$.
    **if** $n \ge 3$ **then**
        Compute $\mu$ and $\sigma$, the mean and the standard deviation of $\mathcal{E}_{n-2}, \mathcal{E}_{n-1}, \mathcal{E}_n$.
        **if** $\sigma/\mu < 0.02$ **then**
            $c :=$ **False**.
        **end if**
    **end if**
    $\beta := 1.2\beta$.
    $T_{check} := \lfloor 1.02T_{check} \rfloor$
**end while**
**return** $\mathbf{\Delta}_n$ and $\mathbf{H}_n$.

Figure 9 depicts the mean of the 10 corrected $D_{zz}(z)$ profiles obtained from each annealing calculations. Notably, the FCE estimator systematically underestimates $D_{zz}$ in the immediate vicinity of the wall, which is consistent with our earlier observations in ref 53. In order to further assess the performance of our correction scheme, we feed the corrected $D_{zz}(z)$ into an SDE integrator and compare the corresponding


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

**Algorithm 4** MHMC($\beta, \mathbf{\Delta}_{\text{in}}, \mathbf{\Sigma}, \delta S, T_{\text{check}}$) - Monte Carlo procedure for simulated annealing

**Procedure** HCMC ($\mathbf{\Delta}_{\text{in}}, \mathbf{\Sigma}, \delta S, \beta, , T_{\text{check}}$).
**Inputs:** An input diffusivity profile $\mathbf{\Delta}_{\text{in}} = \{\mathbf{D}_{\text{in}}(\mathbf{r}_i)\}_{i=1}^{n_b}$, a list of standard deviation matrices $\mathbf{\Sigma} = \{\boldsymbol{\sigma}(\mathbf{r}_i)\}_{i=1}^{n_b}$, an initial step size $\delta S$, an inverse temperature $\beta$, and $T_{\text{check}}$, the number of steps at which early termination condition is checked (initially set to 100).
**Outputs:** An updated diffusivity profile $\mathbf{\Delta}_{\text{out}} = \{\mathbf{D}_{\text{out}}(\mathbf{r}_i)\}_{i=1}^{n_b}$, the final binwise relative entropy profile $\mathbf{H} = \{H_i\}_{i=1}^{n_b}$, the acceptance ratio $a$, the scaling factor $b$.
**Algorithm parameters:**

- $\lambda$, the relative strength of the tension term, with a default value of 0.01.
- $C_{\text{upper}}$ and $C_{\text{lower}}$, the highest and lowest acceptable cumulative acceptance probability, with default values of 0.4 and 0.15, respectively.
- $r_{\text{increase}} = h(\sqrt{0.17/a})$ and $r_{\text{decrease}} = h(\sqrt{0.4/a})$, the ratio by which to increase or decrease $b$ when $a$ crosses $C_{\text{lower}}$ and $C_{\text{upper}}$, respectively. Here, $h(x) = \max(1, \min(2, x))$.
- $\boldsymbol{I}_{\text{indices}}$, the list of bins at which tension is applied.
- $T_{\text{change}} = 5$.
- $N_{\text{iter}} = 10T_{\text{check}}$.

$n := 0$.
Set $\mathbf{\Delta} := \mathbf{\Delta}_{\text{in}}$, $N_{\text{attempt}} := 0$ and $N_{\text{accept}} := 0$.
Initialize $\overline{\mathbf{w}} := \{\overline{w}_i\}_{i=1}^{n_b}$, the binwise weights for the tension term as follows. For bins included in $\boldsymbol{I}_{\text{indices}}$, set $\overline{w}_i := [\boldsymbol{\sigma}_i : \boldsymbol{\sigma}_i]^{-1}$ where ':' corresponds to full contraction of $\boldsymbol{\sigma}_i$ with itself. Otherwise, set $\overline{w}_i := 0$.
Set $b := \delta S$, and compute the initial binwise relative entropy profile $\mathbf{H}$. Set the initial acceptance rate to $a = 0$ and compute the initial energy $\mathcal{U}_0$ and tension $\mathcal{T}_0$ using the expressions given by (22) and (23).
Let $\mathcal{E}_0 := \mathcal{U}_0 + \lambda\mathcal{T}_0$.
**for** $i = 1, 2, \cdots, N_{\text{iter}}$ **do**
  Let $\mathbf{\Delta}_{\text{old}} := \mathbf{\Delta}$, $\mathcal{U}_{\text{old}} := \mathcal{U}_{i-1}$, $\mathcal{T}_{\text{old}} := \mathcal{T}_{i-1}$ and $\mathcal{E}_{\text{old}} := \mathcal{E}_{i-1}$.
  Generate a trial move $\delta\mathbf{\Delta}$ by applying a low-pass filter to $\mathbf{\Omega} = b \times \text{TrialMove}(\mathbf{\Sigma}, \mathbf{H})$ (e.g., via convoluting it with $\mathcal{N}(\mathbf{0}, 2w_b^2\mathbf{I})$) where $w_b$ is the width of each bin. Let $\mathbf{\Delta}_{\text{new}} := \mathbf{\Delta}_{\text{old}} + \delta\mathbf{\Delta}$.
  Compute the energy and tension of the new state, namely $\mathcal{U}_{\text{new}}$ and $\mathcal{T}_{\text{new}}$, using (22) and (23). Let $\mathcal{E}_{\text{new}} = \mathcal{U}_{\text{new}} + \lambda\mathcal{T}_{\text{new}}$.
  Generate a random number $r \sim U(0, 1)$ where $U(a, b)$ is the uniform distribution on the interval $[a, b]$.
  $N_{\text{attempt}} := N_{\text{attempt}} + 1$.
  **if** $r < \exp[-\beta(\mathcal{E}_{\text{new}} - \mathcal{E}_{\text{old}})]$ **then**
    Accept the move and set $\mathbf{\Delta} := \mathbf{\Delta}_{\text{new}}$, $\mathcal{U}_i := \mathcal{U}_{\text{new}}$, $\mathcal{T}_i := \mathcal{T}_{\text{new}}$, $\mathcal{E}_i := \mathcal{E}_{\text{new}}$ and $N_{\text{accept}} := N_{\text{accept}} + 1$.
  **else**
    Reject the move and set $\mathbf{\Delta} := \mathbf{\Delta}_{\text{old}}$, $\mathcal{U}_i := \mathcal{U}_{\text{old}}$, $\mathcal{T}_i := \mathcal{T}_{\text{old}}$ and $\mathcal{E}_i := \mathcal{E}_{\text{old}}$.
  **end if**
  **if** $i > 20$ **and** $\text{mod}(i, T_{\text{change}}) = 0$ **then**
    $a := N_{\text{accept}}/N_{\text{attempt}}$.
    **if** $a > C_{\text{upper}}$ **then** $\delta S := \delta S \cdot r_{\text{decrease}}$.
    **end if**
    **if** $a > C_{\text{lower}}$ **then** $\delta S := \delta S/r_{\text{increase}}$.
    **end if**
  **end if**
  **if** $\text{mod}(i, T_{\text{check}}) = 0$ **and** $\text{Terminate}(\{\mathcal{E}_j\}_{j=1}^{i}, T_{\text{check}}, n)$ **then**
    **return** $\mathbf{\Delta}$, $\mathbf{H}$, $a$ and $b$.
  **end if**
**end for**
**return** $\mathbf{\Delta}$, $\mathbf{H}$, $a$ and $b$.

full and composite van Hove correlation functions with those obtained from MD. As can be seen in Figure 7, annealing considerably improves the agreement between SDE and MD, irrespective of whether fully stochastic or composite van Hove correlation functions are utilized for comparison.

We emphasize that the minor discrepancies that persist even after annealing do not indicate the inadequacy of the annealing process. In fact, the annealing optimization approach works as intended, as demonstrated by the near-disappearance of disagreements when computing the composite $\tilde{G}_s$ values of Figure 7 from diffusion maps, which constitute the basis of the optimization in the first place. For instance, there is a small but significant discrepancy between composite $\tilde{G}_s$ obtained from SDE and $\tilde{G}_s^{\text{MD}}$ at the first peak in density (Figure 10A). The discrepancy, however, almost disappears when diffusion maps are employed. The bin-wise relative entropy is also considerably smaller when a comparison is made based on diffusion maps (Figure 10B). One one hand, diffusion maps


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

**Algorithm 5** Terminate($\{E_j\}_{j=1}^{s}, b, n$) - Early termination condition for the MHMC algorithm

**Function** Terminate($\{E_j\}_{j=1}^{s}, b, n$)
**Inputs**: $\{E_j\}_{j=1}^{s}$ is the total energy time series, and $s$ and $b$ are the total number of steps and block size, respectively.
Let $q := s$ **div** $b$.
$p :=$ **False**.
**if** $q = 1$ **then**
    **return False**.
**else**
    Compute the means and standard deviations of the last two blocks as:

$$\mu_p = \frac{1}{b}\sum_{j=0}^{b-1} E_{s-b-j} \qquad \sigma_p = \sqrt{\frac{1}{b-1}\sum_{j=0}^{b-1}\left(E_{s-b-j} - \mu_p\right)^2}$$
$$\mu_c = \frac{1}{b}\sum_{j=0}^{b-1} E_{s-j} \qquad \sigma_c = \sqrt{\frac{1}{b-1}\sum_{j=0}^{b-1}\left(E_{s-j} - \mu_c\right)^2}.$$

    **if** $|\mu_c - \mu_p| < 2\min\{\sigma_p, \sigma_c\}$ **then**
        $p :=$ **True**.
        $n := n + 1$.
    **else**
        $p :=$ **False**.
        $n := 0$.
    **end if**
    Compute $a$, the slope of a linear regression between the arrays $X = \{1, 2, \cdots, 2b\}$ and $Y = \{E_{s-2b+1}, \cdots, E_s\}$.
    **if** $|a| \leq \tan\frac{\pi}{6}$ **then**
        $c :=$ **True**.
    **end if**
**end if**
**return** $(c \wedge p) \vee (n > 2)$.

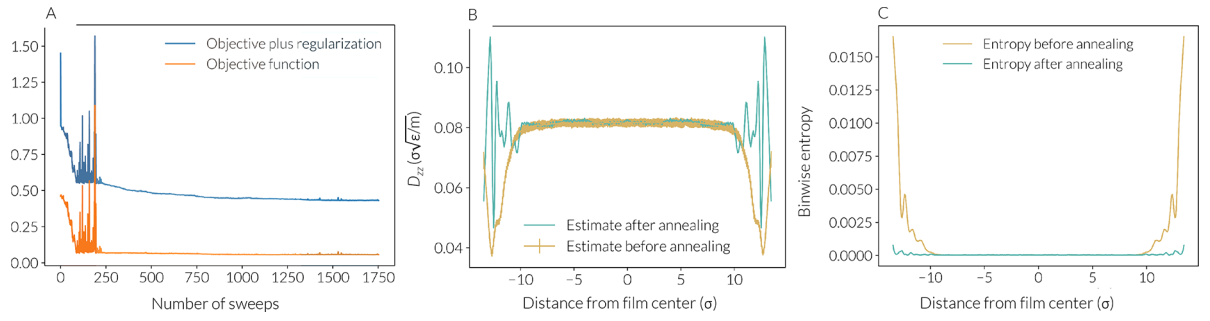



**Figure 8.** Illustration of the performance of the annealing algorithm applied to a $D_{zz}(z)$ profile computed from a single MD trajectory. (A) Value of the objective function (without and with regularization) as a function of MHMC sweeps. The sweeps corresponding to each temperature are depicted using different color. (B) Corrected $D_{zz}(z)$ profile at the end of the annealing. (C) Binwise relative entropy before and after annealing.

offer only an approximation of the Smoluchowski operator and are exact only when $\tau \to 0$. Using them is therefore a trade-off between efficiency and accuracy. Meanwhile, the numerical integration of eq 8 is not exact and is instead subject to uncertainties not only due to sensitivity to implementation details such as the employed time step but also because of the ambiguity in the form and magnitude of the drift term applied in the immediate vicinity of the wall. One can also not rule out the possibility that particle displacements might fundamentally deviate from the canonical diffusive behavior in the immediate vicinity of the wall, in which case eq 7 will fail to capture the statistics of particle mobility. As such, any comparison with SDE-generated data will be irrelevant. Given these possibilities, it is unclear whether the minor discrepancies observed in Figure 7 can ever be fully resolved through improving the optimization algorithm.

### III.D. Comparison with *ad hoc* Approaches.

As discussed in Section I, several authors have employed *ad hoc* definitions of the local MSD to estimate diffusivity. Assuming one-dimensional confinement along the $z$ dimension, such extensions generally take the following form

$$\mathcal{M}(z, t) = \langle |\mathbf{r}_{\parallel}(t + \tau) - \mathbf{r}_{\parallel}(\tau)|^2 \, w(z_\tau, z_{t+\tau}; z)\rangle_\tau \tag{25}$$

where $\mathbf{r}_{\parallel}$ corresponds to the projection of particle position onto the $xy$ plane and $w(z_\tau, z_{t+\tau}; z)$ is a weight function that specifies the contribution of each particle to the local MSD. $D_{\parallel}(z)$ is then estimated from the slope of a linear fit between $t$ and $\mathcal{M}(z, t)$.

In order to compare our FCE estimator with such alternatives, we compute two versions of local MSD using the following weight functions:


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

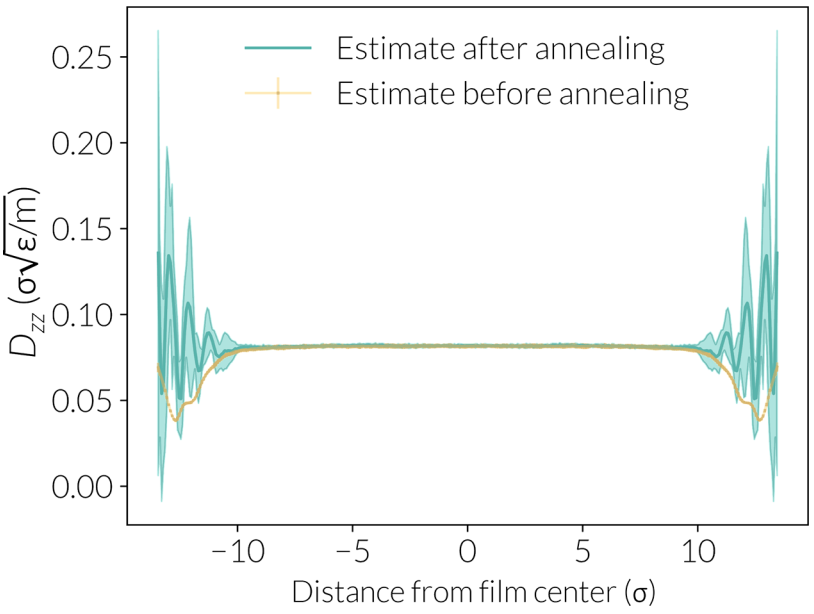



**Figure 9.** Average normal diffusivity profile $D_{zz}(z)$ before and after annealing.

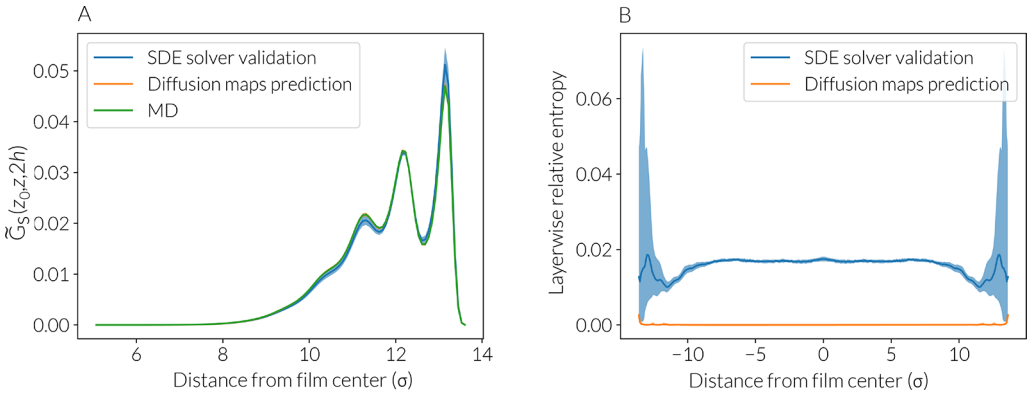



**Figure 10.** (A) Composite $\tilde{G}_s(z, z_0, h, h)$ values and (B) bin-wise relative entropies computed from SDE integrated trajectories and diffusion maps using the annealed $D_{zz}(z)$ profile. For plots in (A), $z_0$ corresponds to the first peak of the number density profile.

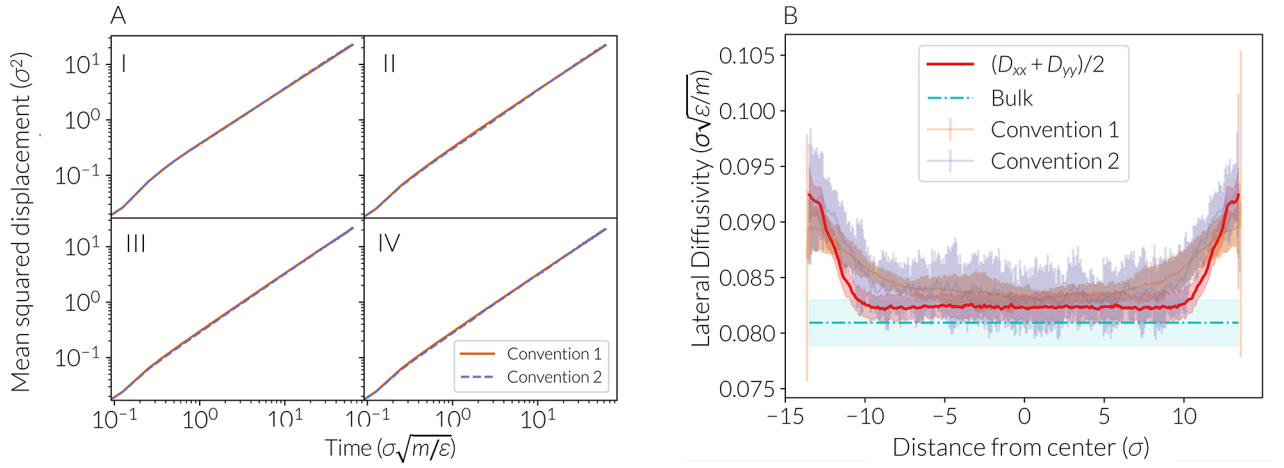



**Figure 11.** (A) *ad hoc* MSDs computed using the conventions given by eqs 26 and 27 at select positions across the film. Labels I–IV corresponds to starting positions highlighted in Figure 3C. (B) Lateral diffusivity profile computed via *ad hoc* MSDs, compared with the FCE estimate obtained using eq 15.

$$w_1(z_\tau, z_{t+\tau}; z) = \frac{\delta(z_\tau - z) + \delta(z_{t+\tau} - z)}{2} \quad (26)$$

$$w_2(z_\tau, z_{t+\tau}; z) = \delta(z_\tau - z)\delta(z_{t+\tau} - z) \quad (27)$$



https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

In the second convention,[32] only particles that reside within the bin both at the beginning and at the end of an observation window contribute to the local MSD of that bin. The first convention, which is inspired by our FCE estimator, includes particles that are within the bin either at the beginning or at the end of the window, giving twice as much weight to those who start and finish at the same bin. Figure 11A demonstrates that the computed MSDs are not very sensitive to the employed convention. The diffusivity profiles computed from such *ad hoc* MSDs (Figure 11B) are also fairly consistent with one another and those obtained using our FCE estimators. It is, however, noteworthy that *ad hoc* profiles appear to be flatter than those obtained from our FCE estimator. The observed flattening of the diffusion profile can be explained by the high likelihood of particles contributing to the local MSD of each bin having visited neighboring bins intermittently. As a result, the diffusivities of these visited bins have an impact on their mobility. In the current system, the lateral diffusivity exhibits minimal changes across the film, rendering this flattening statistically insignificant. Conversely, in systems with large diffusivity changes, this effect can be more pronounced.

## IV. CONCLUSIONS

In this work, we have successfully applied and validated the filtered covariance estimator (FCE) introduced in our previous paper[53] to extract position-dependent and anisotropic diffusivity profiles from MD trajectories. We choose as our observation window the characteristic relaxation time scale of a position-dependent cage escape autocorrelation function. The proposed estimator accurately reproduces diffusivity values computed from the MSD for a bulk LJ liquid. As a simple but nontrivial test case, we assess the performance of the estimator for an LJ liquid confined within a slit pore. The estimator accurately captures the diffusivity tensor's isotropy at the center of the pore, and the estimated diffusivities match with those obtained from bulk simulations at the same number density. The estimator, however, underestimates the normal component of diffusivity in the immediate vicinity of the confining walls, as expected based on our observations in ref 53. To address this, we have proposed an optimization procedure based on simulated annealing and diffusion maps that effectively corrects this systematic bias. The overall performance of the estimator is further validated through a comparison of the self-parts of the lateral and normal van Hove correlation functions computed from MD and estimated by using the diffusivity profile obtained from FCE.

It is important to note that although the annealing procedure leads to a substantial reduction in the discrepancy between the van Hove correlation functions generated by MD and SDE, it does not eliminate it entirely. As outlined in Section III.C, this outcome is not unexpected since the optimization algorithm terminates at a small but nonzero temperature when the change in energy function becomes sufficiently slow. Further rounds of annealing could produce improved minima. However, there are several aspects of the annealing process that require optimization, including the annealing protocol, trial move generation, and energy function. We choose to use simulated annealing for its simplicity and popularity within the molecular simulations community, but it is generally possible to use any other optimization procedure to solve it (eq 20). More efficient optimization schemes to couple with the diffusion maps procedure may be the subject of future work.

While it is generally possible to use any other optimization procedure to solve this problem (eq 20), we find several advantages for using simulated annealing. In addition to its simplicity and popularity within the molecular simulations community, simulated annealing might be able to avoid the "curse of dimensionality", which is a common issue encountered by many deterministic optimization algorithms. This phenomenon arises when the rate of convergence to a local minimum becomes slower as the dimensionality of the problem increases. In our case, the dimensionality of the optimization problem is determined by the number of spatial bins required for a given level of spatial resolution. Hence, deterministic optimization algorithms may perform better when fine discretization of the simulation domain is not required.

Another relevant issue is regularization. The optimization problem (eq 20) treats diffusivity values at different positions independently, disregarding the expectation that diffusivity needs to be smooth functions of position within a single phase. Brute optimization of (eq 20) can therefore lead to overfitting and spurious oscillations in diffusivity, especially given the uncertainties in van Hove correlation functions obtained from MD. It is therefore natural to constrain the optimization problem (eq 20) in such a way that the only admissible solutions are smooth. For that purpose, we introduce the notion of tension given by eq 23, which plays much the same role as a Bayesian prior distribution in the work of Hummer.[44] There are, however, several other constructions that could enforce smoothness. While our method is not necessarily the most efficient or physically natural, it successfully avoids overfitting and ensures smoothness of the estimated diffusivity profiles.

It is worth noting that while our proposed estimator and correction scheme have been validated using a simple geometry, they can be readily extended to more complex systems. Therefore, the application of our methods is not limited to simple or one-dimensional geometries.

Beyond their use for characterizing spatial variations in diffusivity, the estimators developed in this work can also be utilized for providing accurate parameters of mean-field models for describing the dynamics of confined systems. Such models can, for instance, be numerically integrated to collect reliable statistics of particle motion over time scales not accessible via conventional MD or even coarse-grained methods such as Langevin dynamics. Such statistics can provide insights into how materials behave over long time scales and how mobility statistics impact their functional properties, such as catalytic activity, optical properties, and self-healing potency.

## APPENDIX A: MAPPING OF THE PROPOSED CORRECTION SCHEME ONTO A BAYESIAN FRAMEWORK

Our correction scheme to mitigate hard wall artifacts can be nicely mapped onto a Bayesian framework if relative entropy is used as a distance function. As discussed earlier, our correction scheme aims at decreasing some notion of distance between $\mathbf{P}$, the discretized transition probability matrix directly computed from MD, and $\mathbf{Q} = f(\mathbf{D})$, the transition matrix inferred from the diffusivity profile $\mathbf{D}$ according to diffusion maps. Assuming the validity of $\mathbf{Q}$, the conditional probability of observing a particle dynamic trajectory (or a collection of trajectories) with a total number of $n_w$ observation windows can be expressed as



https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659

$$P(\text{data}|\mathbf{Q}) = \prod_{i,j=1}^{n_b} Q_{ij}^{P_{ij} n_w} \tag{A1}$$

Assuming an uninformative prior over $\mathbf{Q}$ (i.e., $P(\mathbf{Q}) \equiv$ const., $P(\mathbf{Q}|\text{data}) \propto P(\text{data}|\mathbf{Q})$), the log of the posterior distribution can thus be expressed as

$$\log P(\mathbf{Q}|\text{data}) = n_w \sum_{i,j=1}^{n_b} P_{ij} \log Q_{ij} + C_1 \tag{A2}$$

where $C_1$ is an additive constant. Since $\mathbf{P}$ also does not depend on $\mathbf{D}$, $n_w \sum_{i,j=1}^{n_b} P_{ij} \log P_{ij}$ will also be constant and can be added to and subtracted from both sides of (eq A2) to conclude

$$\log P(\mathbf{Q}|\text{data}) = -n_w \sum_{i,j=1}^{n_b} P_{ij} \log \frac{P_{ij}}{Q_{ij}} + C_2 \tag{A3}$$

where $C_2$ is a different additive constant. Note that the right-hand sides of eqs A3 and 22 are similar. More precisely, conducting MHMC sampling of the energy function given by eq 22 using an inverse temperature $\beta := n_w n_b$ will enable us to sample from the posterior distribution $\mathbf{Q}$, and minimizing the said objective function will correspond to obtaining the maximum likelihood estimator (MLE) of $\mathbf{Q}$. If $\mathbf{Q}$ can be sampled arbitrarily from the set of all stochastic matrices, then the minimizer will be trivially given by $\mathbf{Q} = \mathbf{P}$. However, if we instead perform the minimization over all choices of $\mathbf{Q}$ such that $\mathbf{Q} = f(\mathbf{D}(\mathbf{r}))$, then this will implicitly define an MLE for $\mathbf{D}(\mathbf{r})$ associated with the mapping $f$ (which is not in general surjective). Moreover, the regularization term given by eq 23 can be seen as $\mathcal{T} = -\log(\mathbb{P}(\mathbf{D}))$ (i.e., as the logarithm of a nonuniform prior on the diffusivity profiles). Putting everything together, we can now express a posterior in terms of diffusivity:

$$-\log P(\mathbf{D}|\text{data}) = n_w \sum_{i,j=1}^{n_b} P_{ij} \log \frac{P_{ij}}{[f(\mathbf{D}(\mathbf{r}))]_{ij}} + \log \mathbb{P}(\mathbf{D}) + C_2 \tag{A4}$$

## AUTHOR INFORMATION


### Corresponding Author

**Amir Haji-Akbari** − *Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520, United States;* orcid.org/0000-0002-2228-6957; Email: amir.hajiakbaribalou@yale.edu

### Authors

**Tiago S. Domingues** − *Department of Chemical and Environmental Engineering, Yale University, New Haven, Connecticut 06520, United States*

**Ronald R. Coifman** − *Department of Mathematics, Yale University, New Haven, Connecticut 06520, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.3c03581


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

A.H.-A. gratefully acknowledges the support from National Science Foundation grants CBET-1751971 (CAREER Award) and CBET-2024473. R.R.C. gratefully acknowledges the support from AFOSR, The Science of Learning from Observations FA9550-21-1-0317. This work was supported as part of the Center for Enhanced Nanofluidic Transport (CENT), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under award no. DE-SC0019112. These calculations were performed at the Yale Center for Research Computing. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant no. ACI-1548562.73.

## REFERENCES


(1) Alcoutlabi, M.; McKenna, G. B. *J. Phys.: Cond. Mater.* **2005**, *17*, R461.
(2) Binder, K.; Horbach, J.; Vink, R.; De Virgiliis, A. *Soft Matter* **2008**, *4*, 1555.
(3) Chaban, V. V.; Prezhdo, V. V.; Prezhdo, O. V. *ACS Nano* **2012**, *6*, 2766.
(4) Chen, J.; Schusteritsch, G.; Pickard, C. J.; Salzmann, C. G.; Michaelides, A. *Phys. Rev. Lett.* **2016**, *116*, 025501.
(5) Chiashi, S.; Saito, Y.; Kato, T.; Konabe, S.; Okada, S.; Yamamoto, T.; Homma, Y. *ACS Nano* **2019**, *13*, 1177.
(6) Forrest, J.; Dalnoki-Veress, K.; Stevens, J.; Dutcher, J. *Phys. Rev. Lett.* **1996**, *77*, 2002.
(7) Swallen, S. F.; Kearns, K. L.; Mapes, M. K.; Kim, Y. S.; McMahon, R. J.; Ediger, M. D.; Wu, T.; Yu, L.; Satija, S. *Science* **2007**, *315*, 353.
(8) Kearns, K. L.; Still, T.; Fytas, G.; Ediger, M. *Adv. Mater.* **2010**, *22*, 39.
(9) Gao, E.; Li, R.; Baughman, R. H. *ACS Nano* **2020**, *14*, 17071.
(10) Fumagalli, L.; Esfandiar, A.; Fabregas, R.; Hu, S.; Ares, P.; Janardanan, A.; Yang, Q.; Radha, B.; Taniguchi, T.; Watanabe, K.; et al. *Science* **2018**, *360*, 1339.
(11) Olivieri, J.-F.; Hynes, J. T.; Laage, D. *J. Phys. Chem. Lett.* **2021**, *12*, 4319.
(12) Gupta, S.; Cochran, H.; Cummings, P. *J. Chem. Phys.* **1997**, *107*, 10335.
(13) Hu, H.-W.; Carson, G. A.; Granick, S. *Phys. Rev. Lett.* **1991**, *66*, 2758.
(14) Pressly, J. F.; Riggleman, R. A.; Winey, K. I. *Macromolecules* **2018**, *51*, 9789.
(15) Zhang, T.; Winey, K. I.; Riggleman, R. A. *Macromolecules* **2019**, *52*, 217.
(16) Malmir, H.; Epsztein, R.; Elimelech, M.; Haji-Akbari, A. *Matter* **2020**, *2*, 735.
(17) Janiak, C.; Scharmann, T. G.; Mason, S. A. *J. Am. Chem. Soc.* **2002**, *124*, 14010.
(18) Johnston, J. C.; Kastelowitz, N.; Molinero, V. *J. Chem. Phys.* **2010**, *133*, 154516.
(19) Haji-Akbari, A.; Debenedetti, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 3316.
(20) Hussain, S.; Haji-Akbari, A. *J. Am. Chem. Soc.* **2021**, *143*, 2272.
(21) Sharma, S.; Debenedetti, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 4365.
(22) Altabet, Y. E.; Haji-Akbari, A.; Debenedetti, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, E2548.
(23) Shaik, V. A.; Peng, Z.; Brady, J. F.; Elfring, G. J. *Soft Matter* **2023**, *19*, 1384.
(24) Carnes, E. C.; Lopez, D. M.; Donegan, N. P.; Cheung, A.; Gresham, H.; Timmins, G. S.; Brinker, C. J. *Nat. Chem. Biol.* **2010**, *6*, 41.
(25) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459.
(26) Toxvaerd, S.; Praestgaard, E. *J. Chem. Phys.* **1977**, *67*, 5291.
(27) Abraham, F. F. *J. Chem. Phys.* **1978**, *68*, 3713.
(28) Subramanian, G.; Davis, H. *Mol. Phys.* **1979**, *38*, 1061.
(29) Magda, J.; Tirrell, M.; Davis, H. *J. Chem. Phys.* **1985**, *83*, 1888.
(30) Liu, P.; Harder, E.; Berne, B. *J. Phys. Chem. B* **2004**, *108*, 6595.

(31) Mittal, J.; Errington, J. R.; Truskett, T. M. *Phys. Rev. Lett.* **2006**, *96*, 177804.
(32) Haji-Akbari, A.; Debenedetti, P. G. *J. Chem. Phys.* **2014**, *141*, 024506.
(33) Haji-Akbari, A.; Debenedetti, P. G. *J. Chem. Phys.* **2015**, *143*, 214501.
(34) Kubo, R. *J. Phys. Soc. Jpn.* **1957**, *12*, 570.
(35) Helfand, E. *Phys. Rev.* **1960**, *119*, 1.
(36) Marrink, S.-J.; Berendsen, H. J. *J. Phys. Chem.* **1994**, *98*, 4155.
(37) Lançon, P.; Batrouni, G.; Lobry, L.; Ostrowsky, N. *Physica A* **2002**, *304*, 65.
(38) Desai, T.; Keblinski, P.; Kumar, S. K. *J. Chem. Phys.* **2005**, *122*, 134910.
(39) Teboul, V.; Simionesco, C. A. *J. Phys.: Cond. Matter* **2002**, *14*, 5699.
(40) Gaskell, T. *J. Phys. C Solid State* **1971**, *4*, 1466.
(41) Frenkel, D., Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, CA, 1996; Vol. *50*.
(42) Fokker, A. D. *Ann. Phys.* **1914**, *348*, 810.
(43) Planck, V. *Sitzber. Preuss. Akad.* **1917**, *24*, 324.
(44) Hummer, G. *New J. Phys.* **2005**, *7*, 34.
(45) Comer, J.; Schulten, K.; Chipot, C. *J. CHem. Theory Comput.* **2014**, *10*, 2710.
(46) Ljubetič, A.; Urbančič, I.; Štrancar, J. *J. Chem. Phys.* **2014**, *140*, 084109.
(47) Chang, P.-W. C. T.; Fok, J. C. *Biophys. J.* **2015**, *109*, 966.
(48) Ghysels, A.; Venable, R. M.; Pastor, R. W.; Hummer, G. *J. CHem. Theory Comput.* **2017**, *13*, 2962.
(49) De Vos, O.; Venable, R. M.; Van Hecke, T.; Hummer, G.; Pastor, R. W.; Ghysels, A. *J. CHem. Theory Comput.* **2018**, *14*, 3811.
(50) Palmer, B. J.; Chun, J.; Morris, J. F.; Mundy, C. J.; Schenter, G. K. *Phys. Rev. E* **2020**, *102*, 022129.
(51) Poggioli, A. R.; Limmer, D. T. *J. Phys. Chem. Lett.* **2021**, *12*, 9060.
(52) Sicard, F.; Koskin, V.; Annibale, A.; Rosta, E. *J. CHem. Theory Comput.* **2021**, *17*, 2022.
(53) Domingues, T. S.; Coifman, R.; Haji-Akbari, A. *J. Phys. Chem. B* **2023**, *127*, 5273.
(54) Comte, F.; Marie, N. *Electron. J. Stat.* **2021**, *15*, 2566.
(55) Marie, N.; Rosier, A. *Scand. J. Stat.* **2022**, 1−49.
(56) Bandi, F. M.; Moloche, G. *Economet. Theor.* **2018**, *34*, 896.
(57) Engel, M. Zenodo 2021, 4639570, DOI: 10.5281/zenodo.4639570.
(58) Nosé, S. *Mol. Phys.* **1984**, *52*, 255.
(59) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
(60) Schuss, Z. *Theory and Applications of Stochastic Processes: An Analytical Approach*; Applied Mathematical Sciences; Springer-Verlag: New York, 2010.
(61) Kob, W.; Andersen, H. C. *Phys. Rev. E* **1995**, *51*, 4626.
(62) Rackauckas, C., Nie, Q. *J. Open Res. Software* **2017**, *5*, DOI: 10.5334/jors.151.
(63) Evans, L.; Cameron, M. K.; Tiwary, P. *J. Chem. Phys.* **2022**, *157*, 214107.
(64) Coifman, R. R.; Lafon, S. *Appl. Comput. Harmon. A* **2006**, *21*, 5.
(65) Trstanova, Z.; Leimkuhler, B.; Lelièvre, T. *P. R. Soc. A- Math. Phys.* **2020**, *476*, 20190036.
(66) Hoerl, A. E.; Kennard, R. W. *Technometrics* **1970**, *12*, 55.


https://doi.org/10.1021/acs.jpcb.3c03581
J. Phys. Chem. B 2023, 127, 8644−8659