


Parallel computing for mobilities in periodic geometries

Martin Magill, Andrew M. Nagel , and Hendrick W. de Haan*

Faculty of Science, University of Ontario Institute of Technology, 2000 Simcoe St N, Oshawa, Ontario L1H7K4, Canada



(Received 12 August 2022; accepted 24 September 2022; published 10 October 2022)

We examine methods for calculating the effective mobilities of molecules driven through periodic geometries in the context of particle-based simulation. The standard formulation of the mobility, based on the long-time limit of the mean drift velocity, is compared to a formulation based on the mean first-passage time of molecules crossing a single period of the system geometry. The equivalence of the two definitions is derived under weaker assumptions than similar conclusions obtained previously, requiring only that the state of the system at subsequent period crossings satisfy the Markov property. Approximate theoretical analyses of the computational costs of estimating these two mobility formulations via particle simulations suggest that the definition based on first-passage times may be substantially better suited to exploiting parallel computation hardware. This claim is investigated numerically on an example system modeling the passage of nanoparticles through the slit-well device. In this case, the traditional mobility formulation is found to perform best when the Péclet number is small, whereas the mean first-passage time formulation is found to converge much more quickly when the Péclet number is moderate or large. The results suggest that, given relatively modest access to modern GPU hardware, this alternative mobility formulation may be an order of magnitude faster than the standard technique for computing effective mobilities of biomolecules through periodic geometries.

DOI: [10.1103/PhysRevE.106.045304](https://doi.org/10.1103/PhysRevE.106.045304)

I. INTRODUCTION

Microfluidic and nanofluidic devices (MNFDs) are an emerging class of biotechnologies with various promising applications in the biological and medical sciences [1–4]. Of these devices, an important subclass of periodic MNFDs exploits the motion of molecules driven through a periodic arrangement of geometric features (e.g., by an electric field) to induce separation by size or other chemical properties. For instance, some of the first MNFDs used for biomolecular separation consisted of periodic arrays of micron-scale posts [5], and work on this type of MNFDs remains an area of active research and development [6–13]. Variants of the post-array design with asymmetric obstacle shapes form the basis of so-called Brownian ratchet devices [14–17]. The slit-well motif is another important MNFD design, consisting of a planar confinement with alternating deeper well regions and shallower slit regions. First pioneered by Han and Craighead [18] and elaborated upon in a series of subsequent studies [19–21], the slit-well device has stimulated ongoing research interest [22–30]. A related MNFD design, the capillary-well motif, has been developed more recently [31,32]. The dynamics of biomolecules have also been studied experimentally, theoretically, and numerically in a variety of other periodic geometries, such as one-, two-, or three-dimensional arrays of spherical cavities [33–37], channels with periodic bands of attractive and repulsive zones on their walls [38–41], a network of interconnected channels named the railroad switch motif [42], planar confinement with an array of nanopits [43],

a series of nanopores connected by microchannels [44], and a periodic sheet of graphene alternating with boron nitride [45]. In fact, even MNFDs whose geometries are uniform in the net direction of motion, such as those used for microcapillary hydrodynamic chromatography [46–49], are (trivially) periodic in this direction.

When periodic MNFDs are used to separate molecules according to size or some other chemical property of interest, this is accomplished by coupling that property to the molecule's net speed through the device. Specifically, the transport rate of analytes is usually characterized by the effective mobility, which is the mean velocity on long timescales normalized by the magnitude of the applied force-generating field. In fact, many molecular mixtures of interest (e.g., DNA, nanoparticles, etc.) exhibit little to no variation in mobility when driven through free solution as the net force and net friction scale in direct proportion to one another. To enable molecular separation, MNFDs break this symmetry by exploiting the interplay of drift and diffusion in nontrivial geometries.

In practice, the design of such devices can be challenging. One aims to control the coupling between mobility and molecular characteristics by optimizing design parameters (such as applied voltage or pressure, solvent composition, and device geometry) to produce the desired profile of effective mobilities. Simulations are often a valuable aid in elucidating the influence of the many design parameters on molecular transport properties such as mobility.

The most common definition of mobility is

$$\mu_{\text{direct}} = \lim_{t \rightarrow \infty} \frac{\langle x(t) \rangle / t}{\Phi}, \quad (1)$$

*Hendrick.deHaan@uoit.ca

where $\langle x(t) \rangle$ is the ensemble mean at time t of the center-of-mass position in the net direction of motion. Here Φ is a scalar characterizing the magnitude of the field that is generating the force driving molecular motion. The choice of Φ is context-dependent: for electrically driven motion, $\Phi \sim \Delta V$ must characterize the gradient of the applied electric potential V (see, e.g., Cheng *et al.* [23]); for pressure-driven motion, $\Phi \sim \Delta p$ should indicate the gradient of the applied pressure p (see, e.g., Ollila *et al.* [27]); and so on. Equation (1) will be referred to here as the direct mobility because it is defined directly in terms of the physical observable it is used to study: the long-term drift velocity of molecules through the system.

In contrast, the focus of this work is the quantity defined as

$$\mu_{\text{indirect}} = \frac{L}{\Phi \langle \tau_1 \rangle}, \quad (2)$$

where L is the period of the system geometry and $\langle \tau_1 \rangle$ denotes the ensemble mean of the first-passage time across one such period. Equation (2) will be referred to as the indirect mobility because it is formulated in terms of observables that can be measured without directly examining the long-term motion across many periods. This manuscript includes a careful comparison of the direct and indirect mobility formulations.

In fact, the two mobility definitions are equivalent under certain circumstances. Indeed, several classical theoretical frameworks imply that $\lim_{t \rightarrow \infty} \langle x(t) \rangle / t = L / \langle \tau_1 \rangle$. However, as reviewed in Appendix A, these results are derived under fairly strong assumptions. Fick-Jacobs theory (Appendix A 1) assumes that motion can be reduced to an effective one-dimensional system, and this approach is typically limited to weakly driven motion and/or slowly varying geometries. Kramers theory (Appendix A 2) and related reaction rate theories assume that most degrees of freedom of the system relax very quickly on the timescale over which the molecule traverses the distance L along the device.

More generally, one can argue for the equivalence of the two mobility definitions based on ergodicity. A single molecule that has crossed a large number k of periods at time t will have sampled the crossing time for a single period k times. The long-time mean of this one particle's k crossing times will be, by ergodicity, equal to the ensemble mean of the time to cross a single period, so that $t \approx k \langle \tau_1 \rangle$. Conversely, its position will be roughly $x \approx kL$, since it has crossed k periods. Thus, its mean velocity will be

$$\frac{x}{t} \approx \frac{kL}{k \langle \tau_1 \rangle} = \frac{L}{\langle \tau_1 \rangle}, \quad (3)$$

from which the equivalence of Eqs. (1) and (2) follow.

A more detailed derivation of this result is included in Appendix B and demonstrated numerically on a test problem in Sec. III A. The equivalence of the two definitions is proven under the simple hypothesis that the system satisfies the Markov property on the timescale of crossing from one period to the next. Specifically, the dynamics of the analyte between the time it first enters the k th period and the time it first enters period $k + 1$ are assumed to depend only on the state of the analyte at the moment that it first entered period k . Under this assumption, the limiting form for the ensemble distribution of x positions can be deduced in closed form. Correlations between the crossing times in consecutive

periods are appropriately taken into account, and these are seen to directly affect the effective diffusion coefficient of the analytes on long timescales.

The equivalence of direct and indirect mobility depends on one crucial technical requirement: the mean first-passage time $\langle \tau_1 \rangle$ in Eq. (2) must be defined with respect to a particular stationary distribution. It is argued in Appendix B 2 that this distribution should exist and be unique under typical conditions. Moreover, a simple Markov chain Monte Carlo algorithm for estimating this distribution numerically is described in Appendix C and tested in Sec. III B.

The limiting behavior of the transport dynamics deduced in Appendix B enables an approximate convergence rate analysis of the two mobility formulations included in Appendix C. It appears that the indirect mobility is better suited for exploiting the massive parallelization afforded by modern hardware. Given reasonable access to such hardware, the analysis suggests that standard computational studies of mobilities through periodic geometries (such as those conducted in Refs. [8,13,22,23,27,37,38,40,41,43,45,49]) may be made to converge up to an order of magnitude more quickly with very little modification to the underlying simulation algorithms. The computational advantage of the indirect mobility in the test case from Sec. II B is verified numerically in Sec. III B.

II. PROBLEM DEFINITION

A. The general case: Transport through a periodic geometry

The physical systems under consideration are those in which a single molecule is driven through a periodic MNFD. The molecular motion is stochastic, such that the physical observables of interest are ensemble averages. Some external force field (e.g., by an applied voltage) biases the stochastic motion of the molecule. The mean direction of the molecule's center-of-mass motion over long timescales will be called the \hat{x} direction. The geometry of the periodic MNFD and the external force field are both taken to be periodic in the direction of \hat{x} , with a period of length L . Every interval of length L in the \hat{x} direction will be called one period of the device.

The molecule traveling through the system will be represented by a finite number of degrees of freedom N_{dof} , which specify all information about the system's state. These would typically be the positions of all the atoms in the molecule. In the event of a time-periodic force field, the phase of the molecule with respect to the period of the force field should also be considered an auxiliary coordinate. In particular, we assume that the dynamics of these N_{dof} degrees of freedom are well-approximated as Markovian, at least on the timescale over which the molecule crosses a period of the device. As reviewed, for instance, by Hänggi *et al.* [50], coarse-grained representations can exhibit non-Markovian dynamics (i.e. memory) even when the underlying system is actually Markovian at the finest scale. Nonetheless, for the models commonly used to study periodic MNFDs, this Markovian assumption is either exactly true or a good approximation on the timescales of interest [8,13,22,23,27,37,38,40,41,43,45,49].

Of the degrees of freedom, the position of the center of mass in the \hat{x} direction at time t will be denoted by the random

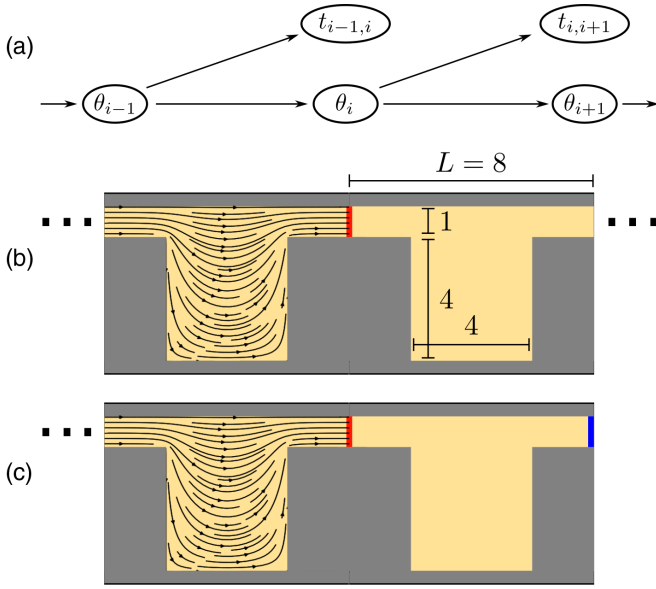


FIG. 1. Schematic illustrating the model of nanoparticles in the slit-well device. (a) A probabilistic graphical model of the Markov chain model from Appendix B. For the problem described in Sec. II B, the only auxiliary coordinate is $\theta_i = y_i$, the y coordinate of the nanoparticle at its first passage to each new period. The distributions of y_{i+1} and $t_{i,i+1}$ are each determined entirely by y_i . (b) In the direct mobility formulation, particles are initialized uniformly in the y direction at the midpoint of a slit (red line) and evolve forward in time until many periods have been crossed. (c) In the indirect mobility formulation, particles are also initialized at the red line and evolve forward in time until a single period has been crossed (blue line). The shape of the baseline electric field E is denoted by the black field lines. The field is periodic in the horizontal direction with period L , such that the field shape is the same in every period of the device. Diffusive motion against the direction of E is possible in both mobility formulations.

variable x or $x(t)$ to make the time dependency explicit. The remaining degrees of freedom will be called auxiliary coordinates and denoted collectively by the random vector θ or $\theta(t)$. The original value of x will be fixed to some $x_0 = 0$ at time $t = 0$, and the thresholds between consecutive periods will be located at $x_i = x_0 + iL$ for each i . The time elapsed between the molecule's first arrival at x_i and its subsequent first arrival at x_{i+1} is denoted $t_{i,i+1}$. The value of the auxiliary coordinates at the moment of first contact with x_i will be denoted θ_i .

The proof in Appendix B is based on a Markov chain model of this general physical scenario. If the underlying molecular dynamics are Markovian, then the system states at first contact with each x_i form a Markov chain. The Markov chain is written in Eq. (B2) and illustrated in Fig. 1(a) as a probabilistic graphical model (i.e., arrows show statistical dependencies). In particular, the value of the auxiliary coordinates θ_i at first contact with x_i entirely determines the distributions of the first-passage time $t_{i,i+1}$ and of the auxiliary coordinates θ_{i+1} at the first passage into the next period.

Based on an application of the Markov chain central limit theorem to this Markov chain, the limiting form of the position distribution $\rho[x(t)]$ at long times t can be deduced. From there, the equivalence of the the direct [Eq. (1)] and indirect

[Eq. (2)] mobility formulations follows readily. However, this use of the Markov chain central limit requires that the initial auxiliary coordinates θ_0 be initialized according to a specific initial distribution. In particular, the initial distribution of auxiliary coordinates $\rho(\theta_0)$ must be chosen such that the auxiliary coordinates θ_1 measured at the first-passage threshold x_1 are distributed according to the same distribution: $\rho(\theta_1) = \rho(\theta_0)$. This choice corresponds to the stationary distribution of the Markov chain in Eq. (B2); Appendix B 2 includes discussion regarding its existence and uniqueness.

In principle, the choice of the initial position x_0 in the above description is arbitrary and should not impact the mobility. In practice, however, shifting x_0 can have subtle but important consequences for the numerical determination of the indirect mobility, as shown in Sec. III B. In particular, x_0 affects the nature of the stationary distribution and thereby controls the computational cost of sampling the initial values of θ for the indirect mobility calculation.

B. Guiding example: Particles in the slit-well device

As a specific illustration of the general circumstance described in Sec. II A, this section presents a model of free-draining nanoparticles traversing the slit-well MNFD under the influence of an applied electric force [18–30]. In particular, we will study the same model analyzed by Cheng *et al.* [23]. Whereas the slit-well has primarily been studied in the context of polymer analytes (especially DNA), we will focus on the more straightforward case of nanoparticle mobilities as it facilitates a more comprehensive numerical exploration. The equivalence of the direct and indirect mobilities is demonstrated numerically for this system in Sec. III A, the task of sampling the correct stationary distribution is explored in Sec. III B, and the computational advantages of the indirect mobility are illustrated in Sec. III C.

The geometry of the system is illustrated in Figs. 1(b) and 1(c). The dimensions are indicated in Fig. 1(b); the period length $L = 8$ and the aspect ratios of the slit and well regions are set to match Cheng *et al.* [23]. The nanoparticles are modeled as hard spherical particles of diameter a having only two degrees of freedom: the x and y coordinates of their centers of mass. The z coordinate of the center of mass is omitted under the symmetry assumption in the z direction, and rotational degrees of freedom are also assumed to be negligible. The applied force field will be held constant in time, so the only auxiliary coordinate, in this case, is $\theta = y$.

Particle motion will be governed by Brownian dynamics, i.e., the overdamped Langevin equation

$$\frac{d\vec{x}}{dt} = -\mu_0 \lambda \nabla U + \sqrt{2D} R(t), \quad (4)$$

where $\vec{x} = (x, y)$, μ_0 is the free-solution mobility of the nanoparticles, D is the free-solution diffusion coefficient, R is a stationary delta-correlated stochastic force with mean 0 and variance 1, λ is a scalar controlling the magnitude of the applied force, and U is the baseline electrostatic potential energy of the particle. The free-solution diffusion coefficients will scale as $D \sim 1/a$ in line with Stokes' law. Following Cheng *et al.* [23], we will focus on the case of free-draining particles: the effective electrostatic force experienced by the

particles is taken to scale with a at the same rate as the friction coefficient ($\gamma = 1/D$), such that the free-solution mobilities remain fixed at $\mu_0 = 1$ for all values of a . The walls are treated as purely reflective conditions applied when the center of the nanoparticles is a distance $a/2$ from the nominal dimensions listed in Fig. 1(b). Hydrodynamic effects are neglected, and electrohydrodynamic phenomena (such as the particle's charge and ζ potential) are subsumed into μ_0 .

The baseline electrostatic potential in Eq. (4) is modeled simply by Laplace's equation,

$$\nabla^2 U = 0. \quad (5)$$

The walls of the slit-well device are treated as perfectly insulating boundary conditions. A unit voltage drop is imposed across one device period measured from the middle of two consecutive slots. By linearity, the field $-\lambda \nabla U$ corresponds to an applied voltage drop of λ per period. To match Cheng *et al.* [23], we define the quantity $E^* = \lambda/L$ as the characteristic field strength. This characteristic field strength E^* is also the correct choice of Φ for computing mobilities in this system [i.e., in Eqs. (1) and (2)]. Finally, it will also be helpful to discuss the system's behavior in terms of the Péclet number $Pé = E^*a$, which is proportional to the drift-diffusion ratio in the system.

This model system's direct and indirect mobilities were computed using particle simulations under various conditions. Equation (4) was discretized using the common Euler-Maruyama scheme [51] to

$$\vec{x}(t_{j+1}) = \vec{x}(t_j) - \mu_0 \lambda \nabla U[\vec{x}(t_j)] \Delta t + \sqrt{2D\Delta t} R_j. \quad (6)$$

Here $\vec{x}(t_j)$ is the position of the particle at time t_j , each R_j is an independent standard normal random variable drawn at each timestep, and Δt is the discrete timestep. A value of $\Delta t = 10^{-3}$ was used for all simulations. The baseline electrostatic potential U and the corresponding baseline field $E = -\nabla U$ were approximated using a mixed finite element method formulation according to the methodology described in Nagel *et al.* [52].

Particle simulations were always initialized with constant x positions but randomly distributed initial y values. The initial x value was generally placed in the middle of a slit [Figs. 1(b) and 1(c)]. The initial values of y were uniformly distributed for the direct mobility calculations. For indirect mobility calculations, the initial values of y were sampled from a precomputed database of 10^5 samples obtained by Markov chain Monte Carlo (MCMC). Specifically, these samples correspond to the final y positions of trajectories initialized with uniform y positions and simulated until a total of $N_{\text{relax}} = 10$ periods were crossed; this ensemble of samples from the stationary distribution was computed once for each choice of E^* and a and reused for all corresponding simulations. A total of 10^7 trajectories were used for indirect mobility calculations. Direct mobilities were computed by simulating 10^4 trajectories until the mean position in the \hat{x} direction exceeded $5000L$, roughly the same methodology used by Cheng *et al.* [23]. The above is the default simulation protocol throughout this paper, but variations of some of these parameters are investigated in Sec. III.

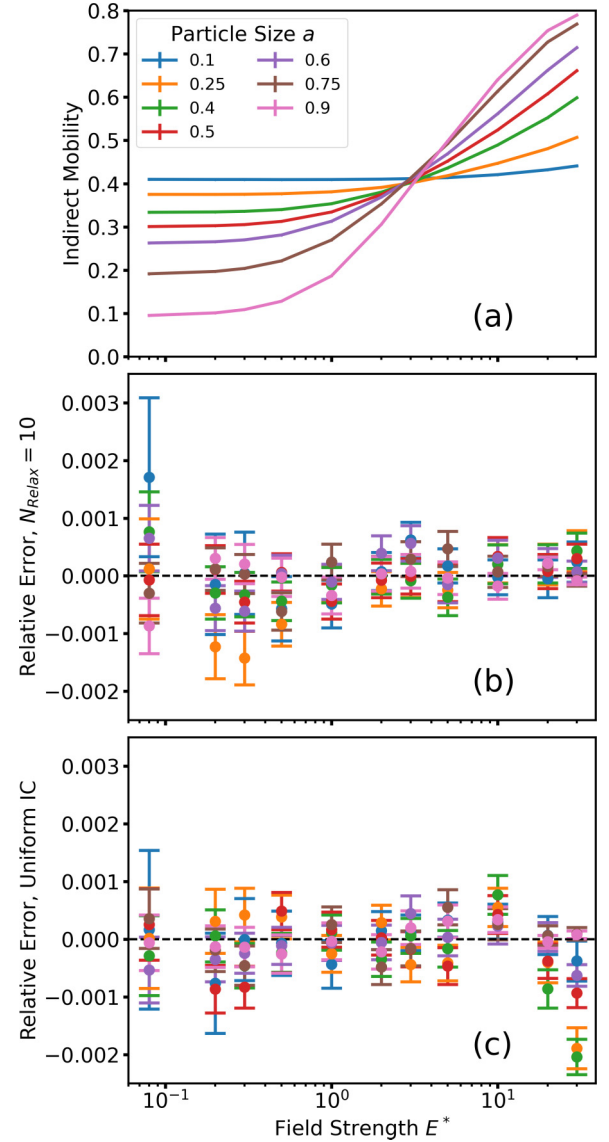


FIG. 2. (a) Measured indirect mobility values as a function of field strength E^* for various particle diameters a . (b), (c) Relative error of simulated indirect mobility values compared to simulated direct mobility values when indirect mobilities are calculated using initial x values in the middle of a slit and (b) initial y values sampled using MCMC with $N_{\text{relax}} = 10$ or (c) initial y values distributed uniformly. Error bars correspond to one standard error.

III. NUMERICAL DEMONSTRATIONS

A. Equivalence of direct and indirect mobilities

Figures 2(a) and 2(b) provides numerical verification that the indirect and direct mobilities are equivalent for the model of free-draining nanoparticles of diameter a through the slit-well MNFD described in Sec. II B. Figure 2(a) shows the indirect mobility as a function of the normalized field strength E^* for nanoparticles of various sizes a . This data can be compared directly to Fig. 2 of Cheng *et al.* [23], where the same measurements were computed using a direct mobility formulation. The direct mobility measurements were also reproduced for the present work, but they are not shown as

they are visually indistinguishable from the indirect mobilities in Fig. 2(a). Instead, Fig. 2(b) shows the relative error of the indirect mobility values relative to the direct mobility measurements. The indirect and direct mobilities were all computed using the default simulation protocol described in Sec. II B. As anticipated, the direct and indirect mobilities are in excellent agreement for all cases in Figs. 2(a) and 2(b): the relative errors are all of the order of 0.1% or better, and all points lie within two standard errors of 0.

B. Sampling the stationary distribution

The results in Figs. 2(a) and 2(b) are based on indirect mobilities calculated by sampling the stationary distribution for y using the MCMC protocol described in Sec. II B. For each choice of E^* and a , particle trajectories are evolved until they traverse $N_{\text{relax}} = 10$ periods, and their final y values form initial conditions for subsequent indirect mobility estimations. That simulation protocol appears sufficient to recover the approximate equivalence of the direct and indirect mobilities. However, because the simulations used for the MCMC algorithm are essentially identical to those used to measure both the direct and indirect mobilities, the MCMC algorithm nominally multiplies the computational cost of the indirect mobility calculation by a factor of N_{relax} . For excessively large values of N_{relax} , any computational advantage of the indirect mobility will be lost.

Luckily, it appears to be possible to reduce this overhead cost dramatically. For example, the protocol in Sec. II B reduces this cost by recycling 10^5 MCMC samples across the 10^7 trajectories used for the indirect mobility calculation. This method reduced the MCMC algorithm's runtime by roughly a factor of 100, rendering it a negligible fraction of the total runtime. The magnitude of the error imparted by this technique will depend on the details of the studied system. Systems for which the true stationary distribution is more intricate and/or for which the first-passage time depends strongly on the initial values of the auxiliary coordinates should incur more error from recycling MCMC samples.

In the current system, however, the stationary distribution for almost all of the physical parameter combinations was found to be very nearly uniform. Figure 3(a) shows histograms of the sampled stationary distributions for all values of E^* and a . The lines are colored according to the drift and diffusion times ratio described below. It is clear that most cases are nearly uniform, and even the few that deviate noticeably from uniform do not deviate very much in absolute terms. Figure 3(b) shows the Kolmogorov-Smirnov test statistics of these distributions with respect to the uniform distribution. This metric is the maximum distance between the empirical cumulative distribution of the MCMC samples of y against the cumulative distribution of a uniform distribution; essentially, a larger value indicates that the samples likely come from a more nonuniform distribution. Here, it is clear that although all of the most nonuniform distributions correspond to large Péclet numbers, not all cases with large Péclet numbers exhibit significantly nonuniform stationary distributions.

In fact, this behavior is not so surprising. The slits of the slit-well device are fairly narrow and long. The local dynamics within each slit likely satisfy the condition assumed in

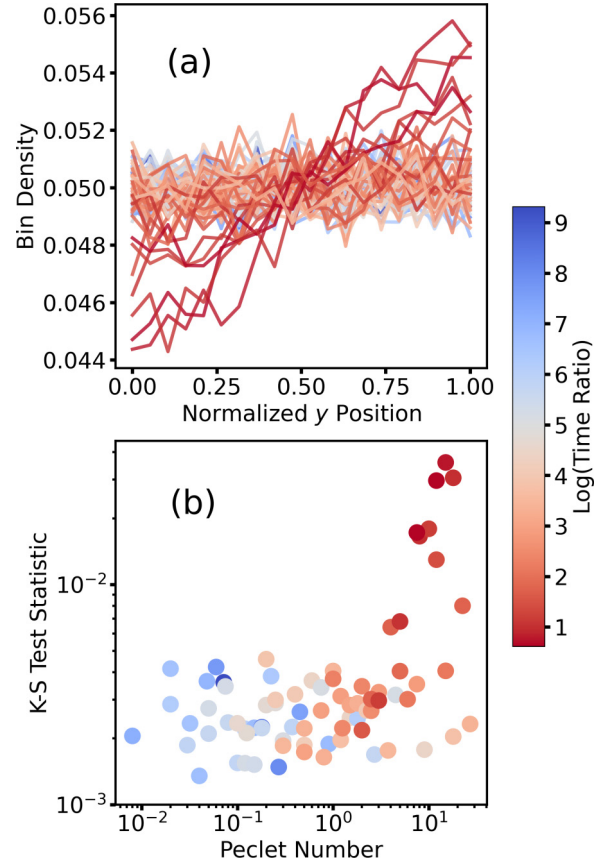


FIG. 3. (a) Normalized histograms (with 20 bins) of the stationary distributions obtained with $N_{\text{relax}} = 10$ for the standard protocol. The normalized y position spans the available y coordinates in the slit, which depends on a through the reflective boundary conditions. (b) Kolmogorov-Smirnov test statistic between the sampled distributions and the uniform distribution (higher value indicates less uniform behavior), as a function of Péclet number. Colors in panels (a) and (b) show $\log(\tau_{\text{drift,slit}}/\tau_{\text{diff,slit}})$ as described in the text.

Fick-Jacobs theory (Appendix A 1): diffusive relaxation of y coordinates occurs much more quickly than translation in the x direction. This behavior explains why not all cases with large Péclet numbers in Fig. 3 exhibit significant Kolmogorov-Smirnov test statistics relative to the uniform distribution: the Péclet number $\text{Pé} = E^*a$ is a global Péclet number and does not account for the local drift-diffusion ratio within the slit, which is more strongly affected by the excluded volume effects due to the particle diameter a .

A local measure of the drift-diffusion balance can be obtained by comparing the drift time along the slit in the x direction,

$$\tau_{\text{drift,slit}} \sim \frac{L/2}{\mu_0 E^*}, \quad (7)$$

to the diffusion time across the slit in the y direction,

$$\tau_{\text{diff,slit}} \sim \frac{(\ell_{\text{slit}} - a)^2}{2D} \sim \frac{a(\ell_{\text{slit}} - a)^2}{2}. \quad (8)$$

Here $L = 8$ is the period length, $\ell_{\text{slit}} = 1$ is the nominal width of the slit, a is the particle diameter, and $D \sim 1/a$ is the particle diffusion coefficient (i.e., the numerical value of D

in simulation was equal to $1/a$). The plots in Figs. 3(a) and 3(b) are colored in proportion to the logarithm of the ratio of these times, $\log(\tau_{\text{drift,slit}}/\tau_{\text{diff,slit}})$. A small ratio corresponds to the Fick-Jacobs regime. Figures 3(a) and 3(b) clearly shows that the most nonuniform stationary distributions are those for which the global Péclet number is high and the drift-diffusion time ratio in the slit is smallest.

In fact, it is even feasible in this case to forego the MCMC sampling step altogether. Figure 2(c) shows relative errors of mobility measurements made with $N_{\text{relax}} = 0$, i.e., a uniform distribution of initial y values. The error relative to the calculated direct mobilities is statistically indistinguishable at all but the largest field strengths. More specifically, as predicted above, based on the ratio of drift to diffusion times in the slit, only for substantial values of E^* and moderate values of a is the error from using a uniform initialization statistically significant in Fig. 2(c). Even at these choices of E^* and a , the indirect mobility still only has errors on the order of 0.02%; the true stationary distributions are still nearly uniform [Fig. 3(a)]. Extending the above reasoning, nearly uniform stationary distributions may be expected to arise in other periodic MNFDs featuring geometric bottlenecks.

Nevertheless, it is important to note that the correct stationary distribution is an essential condition for the equivalence of the direct and indirect mobilities (Appendix B). Figure 4 illustrates the consequences if this condition is neglected inappropriately. In this case, indirect mobilities were once again measured using uniform initial conditions for y , but now with the mean first-passage time computed from an initial x position set in the middle of a well to the middle of the next well (Fig. 5), rather than from the middle of a slit to the middle of the next slit [Fig. 1(c)].

Figure 4(a) shows the indirect mobilities computed based on the well-to-well mean first-passage process with uniform initial conditions, and Figure 4(b) shows the corresponding relative errors. At low field strengths, this algorithm still produces acceptably small relative errors. However, the indirect mobilities are entirely incorrect at higher field strengths, both quantitatively and qualitatively. This behavior is in stark contrast to the results of Fig. 2(c), which showed that uniform initial conditions were an acceptable approximation for all cases in the slit-to-slit configuration.

Indeed, the correct stationary distribution in the well-to-well configuration is substantially nonuniform. Figure 6(a) shows all the distributions for the well-to-well configuration measured with $N_{\text{relax}} = 1$, and Fig. 6(b) shows the corresponding Kolmogorov-Smirnov test statistics relative to the uniform distribution. As in Fig. 3, colors are based on the ratio of the drift timescale to the diffusion timescale; in the well-to-well configuration, these are

$$\tau_{\text{drift,well}} \sim \frac{L/2}{\mu_0 E^*}, \quad (9)$$

$$\tau_{\text{diff,well}} \sim \frac{a(\ell_{\text{well}} - a)^2}{2}. \quad (10)$$

In particular, note that the nominal size of the well is $\ell_{\text{well}} = 5$, which is much larger than the nominal size of the slit, $\ell_{\text{slit}} = 1$. Contrasting with the stationary distributions in the slit-to-slit configuration (Fig. 3), the well-to-well distributions

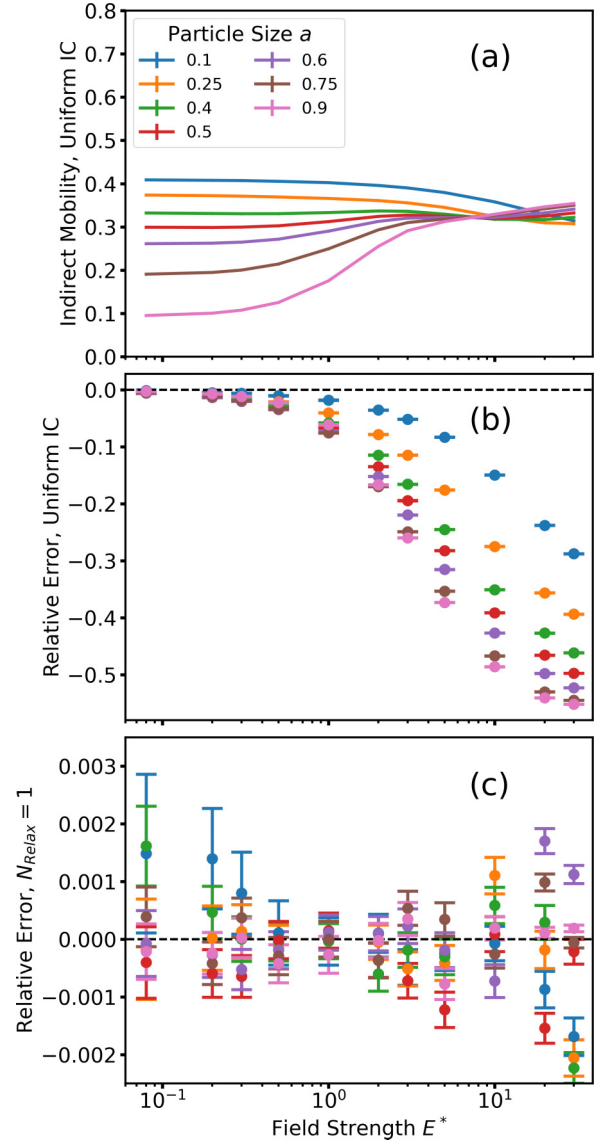


FIG. 4. (a) Incorrect indirect mobility values measured using the well-to-well configuration with uniform initial conditions for y shown as a function of field strength E^* for various particle diameters a . (b), (c) Relative error of simulated indirect mobility values compared to simulated direct mobility values when indirect mobilities are calculated using initial x values in the middle of a well and (b) initial y values are distributed uniformly or (c) initial y values are sampled using MCMC with $N_{\text{relax}} = 1$. Error bars correspond to one standard error.

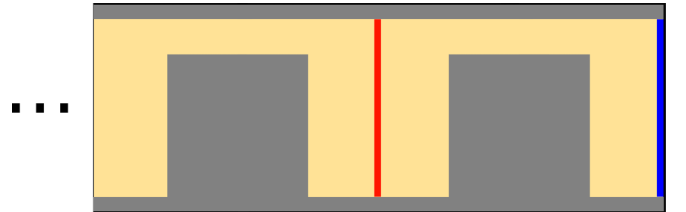


FIG. 5. Schematic of the well-to-well configuration for indirect mobility measurements. Particles are initialized on the red line, which is in the middle of a well. Mean first-passage times are computed to the blue line.

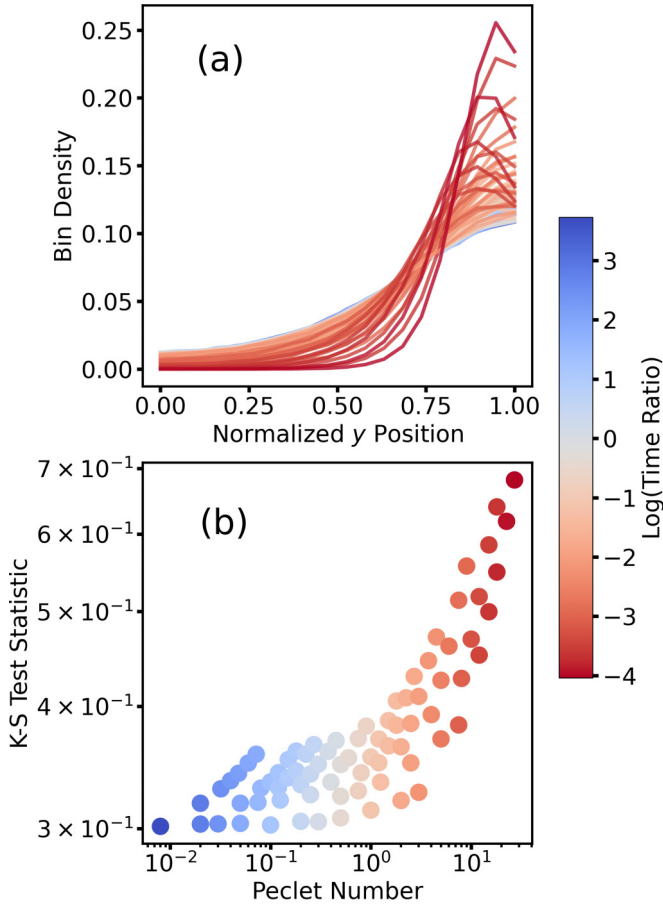


FIG. 6. (a) Normalized histograms (with 20 bins) of the stationary distributions obtained with $N_{\text{relax}} = 1$ for the well-to-well configuration. Normalized y position spans the available y coordinates in the well, which depend on a via the reflective boundary conditions. (b) Kolmogorov-Smirnov test statistic between the sampled distributions and the uniform distribution (higher value indicates less uniform behavior), as a function of Péclet number. Colors in panels (a) and (b) show $\log(\tau_{\text{drift, well}} / \tau_{\text{diff, well}})$ as described in the text.

are clearly far less uniform. With increasing Péclet number, the stationary distribution favours y positions near the top of the well. This behavior is consistent with the known physics of the system: when the applied field is strong, the larger particles are less likely to diffuse into the bottom of the well and have larger mobilities as a result [23].

As in the slit-to-slit configuration, nonuniformity is greatest when the Péclet number is large, and the diffusion timescale in y is large relative to the transit timescale in x. However, whereas the geometric bottleneck in the slit limits the importance of nonuniformity for large a , in the well-to-well configuration, the accessible range in y is large even at the most significant a values. This result highlights a subtle but important aspect of utilizing the indirect mobility formulation in practice: the behavior depends on the choice of location for the period-to-period threshold. This choice affects the complexity of the true stationary distribution and thus affects the possibility of approximating it analytically (e.g., using a uniform distribution) or the computational cost of sampling it numerically. Moreover, the computational disadvantage of

a misaligned period boundary condition is expected to be greatly amplified in systems with larger well-to-slit aspect ratios.

Nonetheless, even the intentionally suboptimal choice of this threshold in the well-to-well configuration can be handled very efficiently in this case. The analysis in Appendix B suggests that the MCMC algorithm should converge exponentially with increasing N_{relax} , suggesting that perhaps large values are not necessary. Figure 4(c) shows the relative error of the well-to-well indirect mobility calculation when the initial conditions are sampled using the MCMC algorithm with $N_{\text{relax}} = 1$. The relative error has become essentially negligible even with relaxation through only a single period. Although errors are statistically discernible at high E^* , these relative errors are on the order of 0.1% or less.

The extremely fast convergence of the MCMC sampling protocol can again be attributed to the geometric bottleneck in the slits. In the well-to-well configuration, the distribution of y values after crossing one period is entirely specified by the intermediate y values in the slit. Because of the bottleneck, the system becomes thoroughly mixed at this location. It thus appears inevitable that the MCMC algorithm will converge very rapidly with N_{relax} whenever such a bottleneck is present.

Future work might explore other options for sampling from the stationary for indirect mobility calculations. Finite samples from the MCMC algorithm presented above might be smoothed by fitting to a histogram or using kernel density estimation, for instance. Alternative, generative modeling techniques (e.g., generative adversarial networks) might be of interest. Furthermore, the indirect mobility can also be applied to solutions of the time-integrated Smoluchowski PDE, as explored in Nagel *et al.* [52] (see Appendix A 3). That application requires estimating the stationary distribution's probability distribution function rather than only requiring samples drawn from that distribution. In the context of the algorithm of Nagel *et al.* [52], this could be accomplished, for instance, by using an auxiliary neural network to represent the stationary distribution and imposing a self-consistency condition on the distribution of y values at the absorbing boundary condition.

C. Computational cost comparison

The demonstration in Sec. III A confirmed that the direct and indirect mobilities are equivalent, as derived in Appendix B. Those simulations used an MCMC algorithm to sample the stationary distribution required for the indirect mobility calculations. As discussed, if every trajectory used to calculate indirect mobility requires an independent MCMC sample, and if the MCMC samples require a large value of N_{relax} , then the cost of the MCMC sampling protocol will dominate the cost of the indirect mobility calculation. Luckily, as explored in Sec. III B, it appears possible to reduce this overhead cost significantly. At least for the model described in Sec. II B, a small number of MCMC samples can be recycled across many trajectories without introducing substantial error. Convergence of the MCMC sampler is expected to be exponential in N_{relax} in general (Appendix B 2), but geometric bottlenecks were argued in Sec. II B to produce particularly fast convergence. Altogether, it appears that the cost of sam-

pling the stationary distribution by MCMC can be made a negligible fraction of the computational cost of the indirect mobility calculation in many cases.

With this established, the current section compares the computational costs of the direct and indirect mobility formulations. It is based on the theoretical cost analysis in Appendix C, which is briefly recounted below. The predicted convergence rates of the two mobilities are tested against their actual performance on the problem of nanoparticles in the slit-well from Sec. II B. The analysis ignores the cost of the MCMC sampling protocol, based on the arguments above that this is likely a small addition to the overall cost of the indirect mobility.

Appendix C contains an analysis of the approximate computational cost of measuring the mobility to a target level of relative error ϵ using either the direct or indirect mobility formulations. The convergence rates are estimated by leveraging the detailed prediction of the limiting x position distribution obtained in Appendix B. Specifically, the limiting distribution is predicted to be a normal distribution with mean and variance given by Eqs. (B15) and (B16), reproduced here for convenience:

$$\langle x(t) \rangle = L \frac{t}{\langle \tau_1 \rangle} + L \frac{1}{2} \frac{\sigma^2}{\langle \tau_1 \rangle^2} - \langle \delta_x \rangle, \quad (11)$$

$$\text{var}(x(t)) = L^2 \frac{\sigma^2}{\langle \tau_1 \rangle^2} \frac{t}{\langle \tau_1 \rangle} + L^2 \frac{5}{4} \frac{\sigma^4}{\langle \tau_1 \rangle^4} + \text{var}(\delta_x). \quad (12)$$

Here L is the period length of the system, $\langle \tau_1 \rangle$ is the mean first-passage time across each period (assuming the stationarity condition for auxiliary coordinates), and σ is a correlation-adjusted standard deviation of the first-passage time across each period [Eq. (B7)].

The quantity δ_x is an additional random variable introduced in Appendix B 4 to account for the motion of analytes against the net long-time direction of motion. In particular, $\langle \delta_x \rangle$ is a measure of the mean fluctuation of the analyte's x position between its first contact with period k and period $k+1$. Because of the stationarity condition, the statistics of δ_x do not depend on k , and thus δ_x plays no role in determining the mobility. However, as discussed below, it does play a very important role in the rate of convergence of the direct mobility calculation.

A measurement of the direct mobility using particle simulations essentially amounts to generating a sample of $x(t)$ values to estimate $\langle x(t) \rangle$ in the direct mobility definition [Eq. (1)]. Appendix C 1 uses the predicted mean and variance of $x(t)$ to deduce the mean relative error [Eq. (C2)] and standard relative error [Eq. (C3)] of the direct mobility estimator in terms of L , $\langle \tau_1 \rangle$, σ , and $\langle \delta_x \rangle$. These errors are expressed as functions of the number of independent trajectories sampled and the total runtime for which the trajectories are evolved. The total relative error of the direct mobility estimator is obtained by adding the mean and standard error in quadrature.

Equation (C2) states that the mean relative error of the direct mobility estimator is proportional to

$$\frac{1}{2} \frac{\sigma^2}{\langle \tau_1 \rangle^2} - \frac{\langle \delta_x \rangle}{L} \quad (13)$$

divided by the total runtime of the simulated trajectories. This prefactor can potentially become very small if its two terms are comparable in magnitude. However, understanding the behavior of δ_x was deemed beyond the scope of the analysis in Appendix C. The direct and indirect mobility convergence rates were compared with the approximation $\delta_x \approx 0$. This choice is one of the significant limitations of that analysis, and the numerical demonstrations below will investigate the implications on convergence rates obtained in practice.

The error convergence of the indirect mobility estimator was approximated by assuming that first-passage times across any given period have exponentially decaying tails. Specifically, the probability density function $\rho(\tau_1)$ of the first-passage time is assumed to be of the form [Eq. (C6)]

$$\rho(\tau_1) \approx \frac{1}{\tau^*} \exp\left(-\frac{\tau_1}{\tau^*}\right) \quad (14)$$

at large τ_1 , where τ^* is some constant. This is generally a fair assumption since the tails of the τ_1 distribution are generated by those stochastic trajectories that remain trapped for extended periods (see the theoretical frameworks in Appendix A 2). However, it is not generally the case that τ^* is equal to the mean first-passage time $\langle \tau_1 \rangle$. Nonetheless, the simplifying assumption $\tau^* \approx \langle \tau_1 \rangle$ was made in parts of Appendix C, as a proper characterization of τ^* is difficult in general. This is the second major limitation in the theoretical comparison between the direct and indirect mobility convergence rates and will also be addressed in the numerical demonstrations below.

The analysis in Appendix C culminates in predictions for the total computation time necessary to achieve a relative error of ϵ using either method when a total of N_{para} parallel threads are available. Figure 7 summarizes the main results of the analysis. Figure 7(a) shows the predicted ratio of the runtimes for the direct and indirect mobility estimators. Results are shown for $N_{\text{para}} = 10^3, 10^4, 10^5$ and with the assumption that the coefficient of variation $\sigma_0/\langle \tau_1 \rangle$ of the first-passage time across a single period is 0.5, 1.0, or 3.0. Here, σ_0 is the actual standard deviation of the first-passage time, which is assumed to be similar to the correlation-adjusted standard deviation σ ; see Appendix C for details.

The general conclusion is that, given sufficient access to parallel computing hardware, the indirect mobility appears to be a more efficient choice. Figure 7(b) illustrates that for target errors below $\epsilon \approx 1/N_{\text{para}}$, the two mobility formulations have roughly the same runtime. Conversely, the maximum advantage of using the indirect mobility occurs for target errors close to $\frac{\sigma_0/\langle \tau_1 \rangle}{\sqrt{N_{\text{para}}}}$, as indicated in Fig. 7(c). As N_{para} increases, the relative cost of the indirect mobility to the direct mobility decreases at all values of ϵ , but the ϵ at which the ratio is maximized shifts to lower values. In practice, ϵ values near 0.1–1% are commonly used in MNFD research, and the N_{para} values listed in Fig. 7 are increasingly affordable thanks to GPU acceleration. Thus, Fig. 7 shows that the theoretical analysis of Appendix C predicts accelerations of an order of magnitude or more by switching to the indirect mobility formulation under practically relevant conditions.

However, as noted above, the analysis in Appendix C and the results in Fig. 7 are based on two questionable

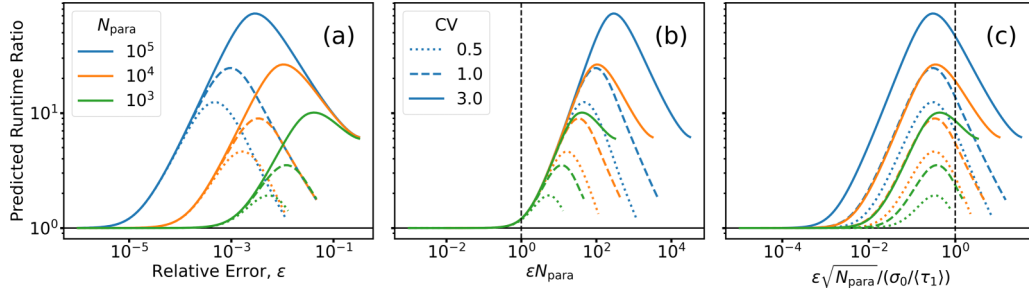


FIG. 7. Ratio of the predicted runtimes for the direct and indirect mobility estimators, for various choices of N_{para} (indicated by line color) and $CV = \sigma_0/\langle\tau_1\rangle$ (indicated by line style). The predicted runtime ratio is plotted against (a) the target relative error ϵ , (b) ϵN_{para} , and (c) $\epsilon\sqrt{N_{\text{para}}} / (\sigma_0/\langle\tau_1\rangle)$. The black vertical line in panel (b) indicates the predicted transition of the direct mobility estimator from order $\mathcal{O}(1/T_{\text{direct}})$ to order $\mathcal{O}(1/\sqrt{T_{\text{direct}}})$, while the black vertical line in (c) indicates the predicted transition of the indirect mobility estimator from exponential convergence to order $\mathcal{O}(1/\sqrt{T_{\text{indirect}}})$. All plots neglect the cost of MCMC sampling and the effect of correlations between first-passage times (i.e., assume $\psi = \sigma/\sigma_0 = 1$; see Appendix C).

approximations. First, it neglects the effect of the quantity $\langle\delta_x\rangle$ in Eq. (C1), which characterizes the motion of analytes against the net force between the first passage to the k th period and the first passage to period $k+1$. Second, it assumes that the first-passage times are exponentially distributed with a time constant τ^* that is similar in magnitude to the mean first-passage time $\langle\tau_1\rangle$. If $\langle\delta_x\rangle$ is large or $\tau^* \gg \langle\tau_1\rangle$, then the predicted computational advantages of the indirect mobility over the direct mobility may be smaller than expected.

Figure 8 presents numerical measurements of the runtimes and estimated relative errors for the direct and indirect mobility estimators, computed from simulations of nanoparticles in the slit-well device (Sec. II B). In each subplot of Fig. 8, four physical scenarios are considered: all combinations of

$E^* = 0.2, 20$ and $a = 0.1, 0.75$. The simulation protocol is the same one used in Sec. III A, with calculations parallelized across $N_{\text{para}} = 10^4$ threads. The direct mobility curves in Figs. 8(a) and 8(c) correspond to measurements taken throughout a single long simulation trajectory. Each data point for the indirect mobility estimators in Figs. 8(b) and 8(d) is sampled independently using a varying number of trajectories (although the same MCMC samples of the stationary distributions are recycled for each physical scenario). Runtimes are reported in simulation time units, and relative errors are estimated against the final direct mobility values for each of the four physical scenarios.

In Figs. 8(a) and 8(b), the colored dotted lines correspond to the predicted runtime necessary to achieve a given relative

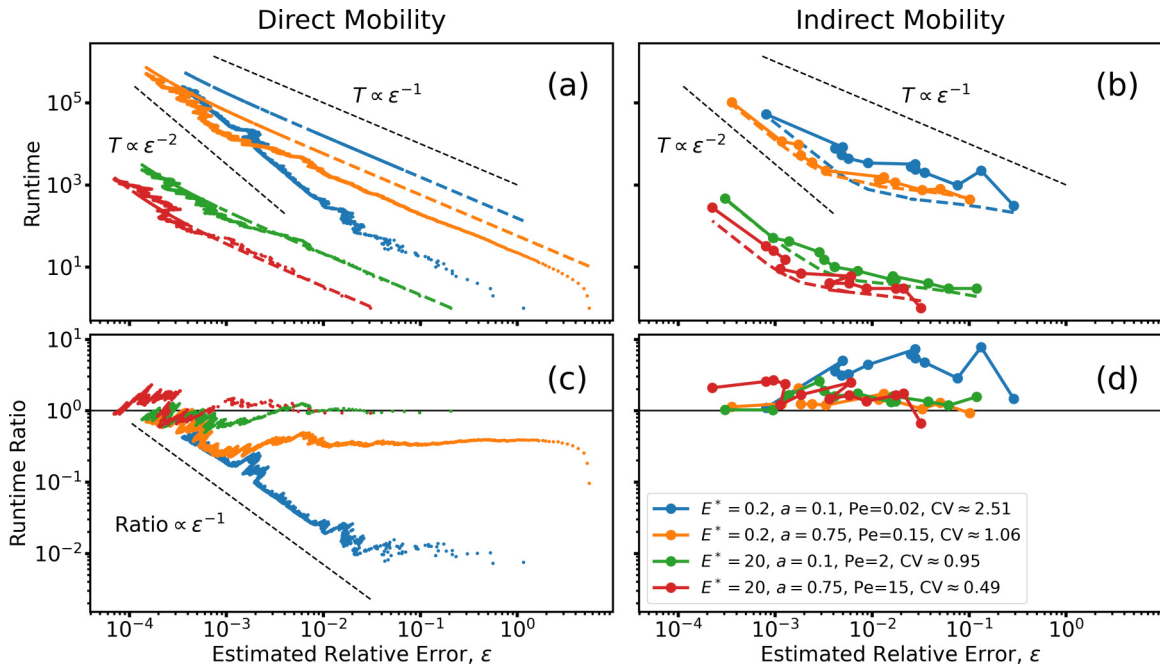


FIG. 8. (a), (b) Measured runtimes for (a) direct mobility estimators and (b) indirect mobility estimators shown as a function of estimated relative error, for four physical scenarios indicated in the legend. Pe indicates the Péclet number $Pe = E^*a$ and CV stands for the coefficients of variation $\sigma_0/\langle\tau_1\rangle$. (c), (d) Deviation of the measured (c) direct mobility and (d) indirect mobility behavior from the predicted behavior. All results for $N_{\text{para}} = 10^4$.

error ϵ [Eqs. (C23) and (C30)]. The black dotted lines denote $T \sim 1/\epsilon$ and $T \sim 1/\epsilon^2$ scaling, corresponding to the expected limiting behavior of the direct mobility estimator. In Figs. 8(c) and 8(d), the measured mobility results are divided by the predicted behaviors; the solid black line indicates where the simulation results and the theoretical predictions are in agreement. The dotted black line indicates a scaling of $1/\epsilon$, as discussed below.

In Fig. 8(a), the red and green lines match the theoretical predictions quite well, but it is clear that the blue and orange series are performing significantly better than expected from the analysis. Specifically, Fig. 8(c) shows that for sufficiently large values of ϵ , the orange line is about twice as fast as expected, and the blue line is roughly two orders of magnitude faster than expected. The deviation by this constant factor persists until a certain level of relative error is achieved, below which the ratio of measured to predicted runtimes decays to 1 at a rate of $1/\epsilon$ [indicated by the dotted line in Fig. 8(c)]. These transition points in Fig. 8(c) coincides with the similar transition points in Fig. 8(a) at which the corresponding runtimes change from scaling as $1/\epsilon$ to scaling as $1/\epsilon^2$.

The disagreement between theory and reality for the convergence of the direct mobility estimator is more prominent for smaller Péclet numbers. This behavior can be tentatively attributed to the omitted term $\langle \delta_x \rangle / L$ acting to reduce the prefactor of the mean relative error by a constant amount. As explained in Appendix C 1, the mean relative error of the direct mobility estimator decays as $\epsilon \sim 1/T$ with runtime T , whereas its standard relative error decays as $\epsilon \sim 1/\sqrt{T}$. At large ϵ , the total error of the direct mobility estimator is dominated by the mean relative error. Since the mean relative error prefactor is missing $\langle \delta_x \rangle / L$, the measured runtime disagrees with the prediction by a constant factor. Conversely, at sufficiently small ϵ , the total error becomes dominated by the standard relative error, which does not depend on $\langle \delta_x \rangle / L$. In the small ϵ regime, the disagreement with the theory decays at the same rate as the mean relative error, i.e., $1/\epsilon$. At large Péclet numbers, however, it appears that indeed $\langle \delta_x \rangle \approx 0$ since the measured runtime versus error agrees well with the prediction from Appendix C.

Whereas the direct mobility is performing much better than predicted in some cases, Figs. 8(b) and 8(d) show that the indirect mobility is performing somewhat less well than expected. The case with the lowest Péclet number (blue) exhibits runtimes nearly an order of magnitude larger than expected over much of the ϵ range. The other three cases (orange, green, red) have runtimes that only exceed the predicted runtime by a factor of 2–3 or less at all values of ϵ . These observations can be attributed to the difference between τ^* and $\langle \tau_1 \rangle$, which were assumed to be equal in the theoretical predictions.

The difference between τ^* and $\langle \tau_1 \rangle$ is better understood by considering the coefficient of variation than the Péclet alone. Indeed, the orange and green lines have very similar coefficients of variation but very different Péclet numbers; whereas only green agrees with theory in the direct mobility case [Figs. 8(a) and 8(c)], both agree comparably well with theory in the indirect mobility case [Figs. 8(b) and 8(d)]. The coefficient of variation is equal to 1 for an exponential distribution, and the case of exponentially distributed first-passage times corresponds to $\tau^* = \langle \tau_1 \rangle$. More generally, the

coefficient of variation of a distribution is a common metric for the relative importance of the distribution's tails. In any case, the magnitude of the gap between theory and practice for the indirect mobility appears less significant than that observed for the direct mobility.

In summary, the theoretical convergence analysis conducted in Appendix C and illustrated in Fig. 7 overpredicts the advantage of the indirect mobility in two ways. When the Péclet number is low, the direct mobility performs better than expected, likely because of the action of δ_x to reduce the mean relative error at large ϵ . When the coefficient of variation is large, the indirect mobility performs worse than expected, likely because τ^* is significantly larger than $\langle \tau_1 \rangle$. Note that in the case of nanoparticles traversing the slit-well device, the Péclet number correlates very strongly with the coefficient of variation in the diffusive regime [23]. Both of these effects tend to diminish the computational advantage of the indirect mobility over the direct mobility.

Regardless, the predicted computational advantage of the indirect mobility is still discernible in this system. Figure 9(b) is a plot of the ratio of the measured runtimes for the direct and indirect mobility estimators shown in Figs. 8(a) and 8(b), obtained by linearly interpolating the direct mobility curves in Fig. 8(a). Also included (dotted lines) are the theoretical predictions of the ratio based on Eqs. (C23) and (C30) (as shown in Fig. 7).

As expected from the discussion of Fig. 8, the measured runtime ratios match the theoretical prediction at high Péclet numbers but significantly deviate at lower Péclet numbers. Nonetheless, the indirect mobility estimator consistently converges faster than the direct mobility estimator for the three largest Péclet numbers for ϵ in the range of 0.1–1%. The largest increase in speed is observed for the green line, which converges roughly six times faster to an error of approximately 0.5%.

Figures 9(a) and 9(c) show how the measured runtime ratios change when these experiments are repeated with fewer parallel threads ($N_{\text{para}} = 10^3$) or more parallel threads ($N_{\text{para}} = 10^5$), respectively. For $N_{\text{para}} = 10^3$, the difference between the two estimators is difficult to resolve at any ϵ value. As noted in Appendix C 3, the two algorithms are expected to have roughly identical convergence rates for small values of N_{para} . In practice, the direct mobility appears slightly more efficient, especially given that this plot omits the cost of sampling the stationary distribution for the indirect mobility estimator.

Conversely, for $N_{\text{para}} = 10^5$, the advantage of the indirect mobility is quite clear [Fig. 9(c)]. In this case, observations are much better described by the theory from Appendix C. Even for the blue line, where the low Péclet number and large coefficient of variation were previously identified as substantially favoring the direct mobility, the indirect mobility calculation is several times faster at errors near 0.1%. The green and red lines, corresponding to the larger Péclet numbers, are well-described by the theory and are 10–20 times faster to compute at errors near 0.1%.

These results demonstrate that using the indirect mobility formulation may indeed be significantly faster under practical conditions. As noted earlier, target errors of 0.1–1% are typically appropriate for simulation studies of periodic MNFDs.

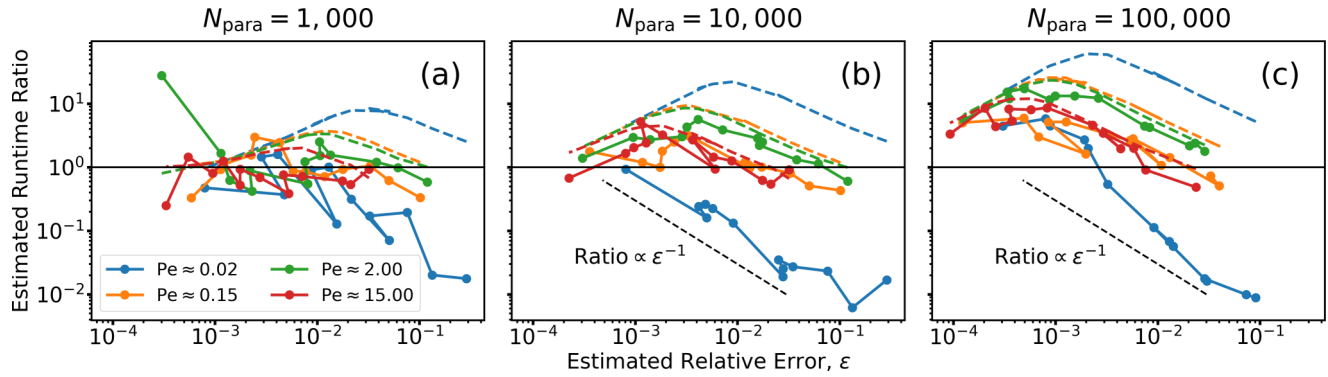


FIG. 9. Ratio of the measured runtimes for direct and indirect mobility estimators as a function of estimated relative error ϵ . Values for the direct mobilities were interpolated/extrapolated linearly to the required values of ϵ . The colored dashed lines show the predicted ratio of runtimes for each case based on Eqs. (C23) and (C30). The solid black line shows a ratio of 1, whereas the dashed black line shows a scaling of $1/\epsilon$. Results shown for (a) $N_{\text{para}} = 10^3$, (b) $N_{\text{para}} = 10^4$, (c) $N_{\text{para}} = 10^5$.

The effective values of N_{para} in practice depend somewhat on implementation details but can be estimated (at least for some implementations on certain NVIDIA GPUs) as 16 times the number of CUDA cores [53]. For our implementation, perfect parallelization to the level of $N_{\text{para}} \approx 10^4$ was easily achieved on a GTX 1650, a very modest consumer-grade GPU with 896 CUDA cores. The case of $N_{\text{para}} = 10^5$ was meant to illustrate what could be achieved by parallelizing across roughly ten such GPUs, a practice which is increasingly becoming commonplace. Experiments on a GTX 1080 Ti suggested that a single card of that variety could deliver $N_{\text{para}} \approx 5 \times 10^4$, consistent with it having 3584 CUDA cores. Although its performance was not tested with our implemented simulations, the RTX 3090 is listed as having 10496 CUDA cores, which may correspond to $N_{\text{para}} \approx 1.5 \times 10^5$ on a single card.

Depending on implementation details, simulations of M -body molecules may effectively be limited to roughly M times smaller values of N_{para} on the same hardware. This would be the case, for instance, if roughly M threads were allocated to accelerate each independent trajectory. This may or may not be advantageous compared to running serial M -body simulations on each of the N_{para} threads; such considerations likely depend on the physics of the system being simulated and are beyond the present scope. Regardless, the increasing affordability of massively parallel computing resources will enable larger N_{para} values for increasingly complex molecules. For instance, a set of 10 RTX 3090 cards could enable $N_{\text{para}} = 10^5$ with M as large as 150, in which case the indirect mobility formulation may feasibly be roughly five times faster than the standard direct mobility formulation, especially for systems with moderate-to-large Péclet numbers and/or geometric bottlenecks.

IV. CONCLUSION

The theoretical and empirical results presented in this work support the claim that the indirect mobility formulation [Eq. (2)] may be a more efficient option for computing the effective mobility of biomolecules driven through periodic geometries than the traditional direct mobility formulation [Eq. (1)]. The indirect mobility formulation leads to exponentially faster convergence in the limit of unlimited parallel

computing capacity and arbitrarily small target errors. Given the growing importance and availability of parallel computing hardware for computational science, the relevance of this result is likely to increase in the future.

Even under realistic conditions of finite parallel computing capacity and target errors near or slightly below 1%, the indirect mobility can still be a substantially more efficient approach. In general, the relative performance of the two approaches appears to depend on a few key physical parameters of the system under study and especially on the balance of drift to diffusion. In the example model of nanoparticles traversing the slit-well device, the indirect mobility was demonstrated to converge up to an order of magnitude faster in some circumstances (specifically, when the Péclet number is moderate or large), even using quite modest computing hardware.

Future work is needed to assess the relative merit of the indirect mobility formulation in simulations of other biophysical systems. The theoretical discussions in the appendices apply to a fairly general model of the transport of biomolecules through periodic geometries. However, the theoretical analyses of computational cost in Appendix C are limited by the approximations of nearly exponential first-passage time distributions and nearly negligible analyte motion against the direction of the applied force. The empirical results reveal that, in some cases, the direct mobility is actually a substantially more efficient estimator than the indirect mobility. Nonetheless, this only appears to be a practical concern in very weakly driven systems. Highly driven systems, where the indirect mobility is most useful, are likely to be of more practical relevance to the design of MNFDs.

ACKNOWLEDGMENTS

The authors gratefully acknowledge funding from Mitacs under the Accelerate Entrepreneur program (Ref. No. IT21168) and from Smart Computing for Innovation (SOSCIP Ref. No. 3-076). H.W.d.H. also acknowledges funding from the Natural Sciences and Engineering Research Council (NSERC) in the form of Discovery Grant No. 2020-07145. M.M. also acknowledges funding from the Ontario Graduate Scholarship (OGS) Program.

APPENDIX A: REVIEW OF RELATED THEORETICAL FRAMEWORKS

Below are included reviews of other theoretical frameworks in which the direct and indirect mobilities have been shown to be equivalent. These derivations are less general than the derivation in Appendix B. The result of Appendix B is more broadly applicable, as it does not make strong assumptions about the separation of timescales in the system, and it holds for many-body molecules and for time-varying force fields. Moreover, Appendix B yields specific equations for the limiting x position distribution, enabling the convergence rate analysis in Appendix C.

1. Quasi-1D systems: Fick-Jacobs theory

The Fick-Jacobs equation was first presented by Jacobs [54] as an effective one-dimensional model for steady-state diffusion in a confined system of varying cross-sections. Essentially, the two- or three-dimensional Smoluchowski equation is integrated across the cross-section of the system. The volume available to particles in these transverse coordinates is approximated in the Fick-Jacobs equation by an additional free energy term. A more rigorous formulation was put forth by Zwanzig [55], who greatly increased the applicability of the equation by formulating a position-dependent effective diffusion coefficient. Further extensions and corrections were proposed in subsequent works [56,57].

The Fick-Jacobs equation and similar approaches have been used successfully to explain diffusion in quasi-one-dimensional systems with or without applied forces, including cases with periodic geometries [58–64]. However, despite the various refinements that have been proposed, it generally fails to perform adequately within certain limits. Because the theory assumes rapid relaxation in the transverse coordinates, it tends to encounter problems in systems with strong applied forces or sudden changes in cross-sectional area [65–72]. Moreover, the theory has a limited capacity to handle spatial variations in the applied force field, especially in the direction of the transverse coordinates (see however Pompa-García and Dagdug [64] for an example where Fick-Jacobs was successfully extended in this manner). Most importantly, for studying biomolecules in periodic MNFDs, the Fick-Jacobs equation is restricted to the diffusion of single Brownian particles and does not directly deal with the case of many-body molecules.

Since the Fick-Jacobs equation is effectively one-dimensional, it benefits from many results applicable to the one-dimensional Smoluchowski equation. While studying the effective diffusion coefficient of one-dimensional Brownian particles in tilted periodic potentials, Reimann *et al.* [73] and Reimann *et al.* [74] proved that the indirect mobility [Eq. (2)] is equal to the direct mobility [Eq. (1)]. Lindner *et al.* [75] connected that work to a classical result due to Stratonovich [76]. These results have since been used to compute mobilities in periodic quasi-one-dimensional systems [58,70,71]. However, this proof of equivalence is naturally restricted to the scope of applicability of Fick-Jacobs theory. Moreover, because it is based on one-dimensional approximations of point-particles, arguments like those in Reimann *et al.* [73] cannot account for correlations between crossing times (cf. Appendix B 2).

2. Large barriers: Poisson point processes

The equivalence of the direct and indirect mobilities is also known to hold in the case that a large free energy barrier obstructs the transport of molecules across each period of the system. In such a setting, there is a well-defined separation of timescales between the period-to-period transport process and all other processes occurring in the system. Reaction rate theories, such as Kramers theory, can be brought to bear on the problem (see Hänggi *et al.* [50] for a review of such theories).

Systems in this regime can be described as Poisson point processes. The probability that a particle initially trapped in a given period has not yet escaped to the next period decays exponentially as a function of time:

$$P(\text{not absorbed after time } t) \sim \exp(-\lambda t). \quad (\text{A1})$$

Moreover, transfers between distinct periods will certainly be statistically independent events since, by assumption, these events occur more slowly than all other relaxation timescales in the system. For such exponentially distributed times, the mean rate λ is related to the mean first-passage time τ by

$$\lambda = \frac{1}{\langle \tau \rangle}. \quad (\text{A2})$$

In the context of mobility through periodic geometries, the mean position on long timescales will thus be

$$\langle x(t) \rangle \rightarrow L\lambda t = \frac{Lt}{\langle \tau \rangle}. \quad (\text{A3})$$

Dividing both sides by t yields the equivalence of Eqs. (1) and (2).

These theories have been used to analyze particle transport in titled periodic potentials [50]. However, Kramers theory and related reaction rate perspectives are restricted in applicability by their strong assumption of timescale separation. Whereas Fick-Jacob methods assume rapid relaxation of position coordinates in the directions transverse to bulk motion, reaction rate theories generally assume rapid relaxation of all processes but the dominant transport process. This approximation again breaks down in situations with strong driving forces and nonequilibrium effects.

Despite their limitations, these theories are still widely used to describe motion in periodic MNFDs. In particular, the assumption of exponentially distributed times is often used to justify physical models based on mean first-passage times (see, for instance, Han *et al.* [19], Cheng *et al.* [23], and Wang *et al.* [41], for a few examples of such arguments). The results presented in this work show that the connection between transport rates and mean first-passage times can be extended to more general physical circumstances, so long as the stationary distribution upon which the mean first-passage times are based is defined appropriately.

3. Other mean first-passage time methods

Besides the Fick-Jacobs and reaction rate theoretical frameworks, there have also been a variety of other cases in which mean first-passage times were used to understand the mobilities of molecules traversing periodic geometries. In particular, mean first-passage time perspectives have been used successfully to study the driven diffusion of

Brownian particles in geometries with abruptly changing cross-sections, where the Fick-Jacobs perspective is not applicable [66–69,72]. We will also mention in passing that these studies have successfully used mean first-passage time frameworks to understand effective diffusion, another important concept in the research and development of periodic MNFDs. The Lifson-Jackson method is one of the earliest methods that studied effective diffusion from this perspective [69,77]. We will briefly comment on how our result connects to the concept of effective diffusion in Appendix B 4, but leave more careful considerations of this aspect for future work.

The Smoluchowski equation describing the evolution of the position probability density function for Brownian molecules is another important theoretical framework for mean first-passage time analysis. An adjoint equation to the Smoluchowski equation can be constructed whose solution at any point in the domain equals the mean first-passage time from that point to absorbing regions on the domain's boundary [78]. In fact, essentially, this method was used by Lifson and Jackson [77] in their analysis.

Another very similar equation is the time-integrated Smoluchowski equation, whose solution is commonly denoted g_0 [79]. The source term in the time-integrated Smoluchowski equation corresponds to a certain choice of initial particle positions in the system, and the solution g_0 has the property that its integral over the domain equals the mean first-passage time. A recursive hierarchy of equations can be constructed to obtain the higher moments of the first-passage time in the same manner.

The qualitative behavior of g_0 solutions in a periodic MNFD was studied by Magill *et al.* [44]. This analysis motivated a certain normalization of first-passage times, which elucidated a universal scaling behavior across system geometries. Moreover, Magill *et al.* [44] argued that the long-time of molecules traversing that MNFD was entirely determined by the first and second moments of the first-passage times across each period, which would be completely captured by the g_0 and g_1 fields.

Elaborating on the g_0 field as a proxy for connecting MNFD geometries with effective mobilities, Nagel *et al.* [52] used a method based on neural networks to solve g_0 in a system similar to that studied by Cheng *et al.* [23]. By computing four-dimensional approximations of g_0 as a function of both domain coordinates and model parameters, Nagel *et al.* [52] demonstrated the idea of using neural networks to construct differentiable mappings from system design parameters to physical observables of interest (in this case, effective mobility). The use of g_0 in this way is a particular motivation for understanding the indirect mobility; the direct mobility formulation cannot be expressed in such a straightforward manner as the solution to a partial differential equation.

APPENDIX B: DERIVATION OF THE EQUIVALENCE OF DIRECT AND INDIRECT MOBILITIES

This section presents a proof that the indirect mobility [Eq. (2)] is equivalent to the more common direct mobility [Eq. (1)], so long as the initial conditions used to compute the indirect mobility are chosen correctly. The approach of the proof is to derive the limiting form of the position distribution

$\rho(x)$ at long times in terms of the mean first-passage time across a single period $\langle\tau_1\rangle$. From this solution, it is possible to equate the limiting drift velocity $\lim_{t\rightarrow\infty}\langle x(t)\rangle/t$ to $L/\langle\tau_1\rangle$. It then follows readily that the two mobility definitions are equivalent.

The general setup for the proof (Appendix B 1) is very similar to the arguments presented previously by Reimann *et al.* [73] and Magill *et al.* [44]. The final steps of the proof (Appendices B 3 and B 4) are very similar to the steps taken by Magill *et al.* [44]. The argument justifying $x \approx kL$ despite analyte backflow (Appendix B 4) is essentially the same used by Reimann *et al.* [73]. However, the first part of the derivation (Appendix B 2) differs substantially from prior derivations to account for correlations in the crossing times between periodic subunits. Such correlations were absent in the system studied by Magill *et al.* [44] because of geometric bottlenecks between the periodic subunits and were irrelevant to the study of Reimann *et al.* [73] which considered only Brownian point particles in a one-dimensional system. They are handled here via the judicious application of the Markov chain central limit theorem to an appropriately constructed Markov chain model of the transport process.

1. The time to first cross k subunits

Recall from Sec. II A that x_i denotes the threshold into the i th period, θ_i denotes the values of the auxiliary coordinates measured at the first time for which $x(t) = x_i$, and $t_{i,i+1}$ denotes the time between first contact with x_i and first contact with x_{i+1} . Now let us denote by τ_k the total first-passage time from the original analyte position at $x = x_0$ to the threshold of the k th periodic subunit at $x = x_k$. By definition,

$$\tau_k = t_{0,1} + t_{1,2} + t_{2,3} + \cdots + t_{k-1,k}, \quad (\text{B1})$$

where $t_{i,i+1}$ is the time to reach x_{i+1} for the first time after having reached x_i for the first time. In the rest of this section, the index k will be used to indicate the total number of channels being crossed, whereas the index i with $0 \leq i \leq k-1$ will be used to refer to the intermediate channels crossed along the way to the k th channel.

Since τ_k is the sum of a series of random variables, it is tempting to appeal to the central limit theorem to deduce its limiting distribution. However, the application of the central limit theorem would require that the random variables $\{t_{i,i+1}\}_{i=0}^{k-1}$ be identically distributed and uncorrelated. As will be shown in Appendix B 2, it is usually possible to initialize the auxiliary coordinates θ_0 such that the $\{t_{i,i+1}\}_{i=0}^{k-1}$ are indeed identically distributed. However, it is not generally possible to eliminate the correlations between the crossing times. Specifically, the correlation of $t_{i,i+1}$ with $t_{i-1,i}$ is mediated by the auxiliary coordinates θ_i measured at first contact with x_i .

Conveniently, the nature of these correlations is still very tractable. The Markovian assumption made in Sec. II A amounts to the statement that the sequence $\{\theta_i\}_{i=0}^{k-1}$ is a Markov chain. Thus, the random process

$$(\theta_0, t_{0,1}) \rightarrow (\theta_1, t_{1,2}) \rightarrow \cdots \rightarrow (\theta_k, t_{k,k+1}) \rightarrow \cdots \quad (\text{B2})$$

is also a Markov chain. Incidentally, since θ_i alone completely specifies the joint distribution of $(\theta_{i+1}, t_{i+1,i+2})$, Eq. (B2) is a special type of Markov chain known as a hidden Markov

model; however, this has no bearing on the current analysis. What is important is that the $t_{i,i+1}$ are fixed observables (i.e., real-valued functionals) of the state $(\theta_i, t_{i,i+1})$.

2. The stationary distribution

The distribution of τ_k can be deduced by applying the Markov chain central limit theorem to the Markov chain given by Eq. (B2). This theorem generalizes the central limit theorem, which applies to a sum of independent and identically distributed (i.i.d.) random variables to the cumulative sum of real-valued functionals of a stationary Markov chain. In particular, we will consider the functional $g(\theta_i, t_{i,i+1}) = t_{i,i+1}$. Note that many variations and extensions of the Markov chain central limit theorem exist, but we only need to appeal to the version of Doob [80].

To apply the theorem, it is necessary for the Markov chain to be initialized in its stationary distribution π , which satisfies

$$\pi(\theta_{i+1}, t_{i+1,i+2}) = \pi(\theta_i, t_{i,i+1}). \quad (\text{B3})$$

In particular, because the marginal distributions of each $t_{i,i+1}$ are completely determined by θ_i , this reduces to the requirement that

$$\pi(\theta_{i+1}) = \pi(\theta_i). \quad (\text{B4})$$

In general, the existence and uniqueness of a Markov chain's stationary distribution π depend on the details of its transition operator. If the state space is finite, then it is sufficient for the transition to be irreducible and aperiodic. This is the case, for instance, when every state θ_{i+1} has a nonzero probability of occurring after any state θ_i . Physically, this type of behavior is common: motion driven by Brownian noise, for instance, usually behaves this way.

Unfortunately, ensuring the existence and uniqueness of a stationary distribution π can be challenging in the case of Markov chains with continuous state spaces—irreducibility and aperiodicity of the transition operators are no longer sufficient conditions. Alas, this is probably the more common scenario in biophysics, for instance, where the auxiliary coordinates θ are the atomic coordinates of a molecule and space is modeled as continuous. Various conditions are known to ensure existence and uniqueness of stationary distributions for continuous state spaces; see for instance Doob [80] or Harris [81]. In particular, if θ_{i+1} is distributed according to a probability density function that is continuous in θ_i , irreducible, and aperiodic, then Eq. (B2) satisfies the conditions of Example 2 on page 215 of Doob [80]. Under these conditions, the stationary distribution exists and is unique, and the Markov chain converges exponentially fast to this stationary distribution from any initial condition. The auxiliary coordinates θ are likely to satisfy this condition (at least to a very good approximation) in most relevant biophysical models.

The exponential convergence of Eq. (B2) to its stationary distribution π suggests that Markov Chain Monte Carlo is a practical method for sampling from π . That is, if Eq. (B2) can be initialized in any convenient state θ_0 and the evolution of the system is simulated until its first passage through k_{relax} of periods through the device, then the final state $\theta_{k_{\text{relax}}}$ will be approximately sampled from π . The number of relaxation periods k_{relax} should not need to be very large if the

convergence of Eq. (B2) to π is indeed exponential for the system under study. The computational cost of this sampling method will be neglected from the cost analysis of computing the indirect mobility in Appendix C. However, an empirical examination of its performance in practice will be presented in Sec. III B for the example of nanoparticles in the slit-well system (Sec. II B).

Finally, assuming that the system is initialized according to the stationary distribution, the Markov chain central limit theorem can be applied to deduce the distribution of Eq. (B1). In general, the Markov chain central limit theorem states that in the limit of large k , for any real-valued function g of the stationary Markov chain state $(\theta_i, t_{i,i+1})$,

$$\rho \left[\sum_{i=0}^{k-1} g(\theta_i, t_{i,i+1}) \right] \rightarrow \mathcal{N}[k\langle g(\theta_0, t_{0,1}) \rangle, k\sigma^2]. \quad (\text{B5})$$

This result closely resembles the classical central limit theorem. For instance, $\langle g(\theta_0, t_{0,1}) \rangle$ is the ensemble average of $g(\theta_0, t_{0,1})$ taken with respect to the stationary distribution π . However, the quantity σ in Eq. (B5) is not simply the variance of g ; see below.

For the choice $g(\theta_i, t_{i,i+1}) = t_{i,i+1}$, and since $\tau_k = \sum_{i=0}^{k-1} t_{i,i+1}$ and $\tau_1 = t_{0,1}$, it follows that

$$\rho(\tau_k) \rightarrow \mathcal{N}(k\langle \tau_1 \rangle, k\sigma^2), \quad (\text{B6})$$

where $\langle \tau_1 \rangle$ is the mean first-passage time across the first periodic subunit when the analytes are initialized according to $\pi(\theta_0)$. The parameter controlling the variance of τ_k is

$$\sigma^2 = \text{var}_\pi(t_{0,1}) + 2 \sum_{i=1}^{\infty} \text{cov}_\pi(t_{0,1}, t_{i,i+1}), \quad (\text{B7})$$

where var_π and cov_π denote variances and covariances, respectively, computed when the system is initialized according to $\pi(\theta_0)$. Since the $t_{i,i+1}$ are all identically distributed, the relationship can be rewritten in the form

$$\sigma^2 = \text{var}_\pi(t_{0,1}) \left[1 + 2 \sum_{i=1}^{\infty} \text{corr}_\pi(t_{0,1}, t_{i,i+1}) \right], \quad (\text{B8})$$

where $\text{corr}_\pi(t_{0,1}, t_{i,i+1})$ are the correlations between distinct crossing times. The first term is the variance of the first-passage time across any single periodic subunit. The terms in the series capture the correlations in the passage times $t_{i,i+1}$ across distinct subunits i , which are mediated by the correlations in the degrees of freedom θ_i . In the special case where these correlations are all zero, we recover i.i.d. behavior in the $\{t_{i,i+1}\}_{i=0}^{k-1}$ and the result reduces to the standard central limit theorem.

3. The number of subunits k that have been crossed at least once at the time t

Consider the (discrete) random variable $\tilde{k}(t)$, the number of channels that the analyte has crossed at least once at time t . The probability that $\tilde{k}(t)$, exceeds some threshold k is given by

$$P(\tilde{k}(t) \geq k) = P(\tau_k \leq t) = \int_0^t \rho(\tau_k) d\tau_k. \quad (\text{B9})$$

Under the conditions leading to Eq. (B6), this integral can be approximated for large k (or, equivalently, large t) as

$$P[\tilde{k}(t) \geq k] \approx \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{t - k\langle\tau_1\rangle}{\sqrt{2k\sigma^2}} \right) \right]. \quad (\text{B10})$$

From Eq. (B10), the probability mass function of k at any (large) time t can be obtained as

$$P[\tilde{k}(t) = k] = P[\tilde{k}(t) \geq k] - P[\tilde{k}(t) \geq k+1]. \quad (\text{B11})$$

However, a more useful form can be deduced by making a discrete-to-continuous approximation, i.e., by pretending that k is a continuous random variable. When k is large, which is the limit of interest, this is an arbitrarily good approximation. Given this, it is therefore sensible to say that the limiting probability density function for k after a long time t is

$$\rho_t(k) \approx -\frac{\partial P[\tilde{k}(t) \geq k]}{\partial k} = \frac{t + k\langle\tau_1\rangle}{\sqrt{8\pi k^3 \sigma^2}} \exp \left[-\frac{(t - k\langle\tau_1\rangle)^2}{2k\sigma^2} \right]. \quad (\text{B12})$$

Although $\rho_t(k)$ is a probability density function, the corresponding probability mass function is approximately

$$P[\tilde{k}(t) = k] \approx \int_{k-0.5}^{k+0.5} \rho_t(k) dk \approx \rho_t(k). \quad (\text{B13})$$

At long times, $\rho_t(k)$ changes very little from $k - 0.5$ to $k + 0.5$, and this approximation is again arbitrarily good. In other words, $\rho_t(k)$ can be interpreted fairly as the probability that, at time t , the analyte has reached x_k at least once but has not yet reached x_{k+1} .

4. The position distribution at long times

Equation (B12) does not directly describe the position of the analyte at a time t . During the time interval after its first passage to x_k and before its first passage to x_{k+1} , the analyte can potentially move to any position with $x < x_{k+1}$. However, as argued below, the distinction between $k(t)$ and $x(t)/L$ is negligible at long times, so Eq. (B12) is in fact an acceptable proxy for the position distribution. The discussion is in the same spirit as that put forth in Reimann *et al.* [73].

Consider the analyte's x position, $x(t)$, in the time interval of duration $t_{k,k+1}$ occurring between τ_k (when it first reaches x_k) and τ_{k+1} (when it first reaches x_{k+1}). Write $x(t) = x_{k+1} - \delta_x(t)$, such that $\delta_x(t)$ is the distance from the analyte's current position to x_k . The motion of the analyte in the direction of $-\hat{x}$ carries an energetic cost, as it opposes the direction of the applied force. This is in addition to any entropic cost incurred for moving through the periodic MNFD. Thus, the probability of observing the analyte at $x = x_{k+1} - \delta_x$ at any point during this time interval will decrease rapidly when $\delta_x \gg L$.

More importantly, by the system's periodicity and the Markov chain's stationarity, it must be that the dynamics of the random variable δ_x do not depend on k (although correlations between consecutive period crossings are possible). Thus, the typical size of δ_x at any time between τ_k and τ_{k+1} is independent of k . For sufficiently large k , the typical distance that an analyte might move in the $-\hat{x}$ direction after reaching x_k and before reaching x_{k+1} is therefore arbitrarily small compared to the total distance it has traveled since $t = 0$. Similarly, the duration $t_{k,k+1}$ will be a small fraction

of the total time τ_k . Thus, although the analyte may briefly move short distances away from x_k before reaching x_{k+1} , these fluctuations do not affect the ensemble dynamics in the limit of long time or, equivalently, large k .

Making the substitution $x \approx kL - \delta_x$ in Eq. (B12) yields

$$\begin{aligned} \rho_t(x + \delta_x) &\approx \frac{Lt + (x + \delta_x)\langle\tau_1\rangle}{\sqrt{8\pi(x + \delta_x)^3(L\sigma^2)}} \\ &\times \exp \left(-\frac{(Lt - (x + \delta_x)\langle\tau_1\rangle)^2}{2(x + \delta_x)(L\sigma^2)} \right). \end{aligned} \quad (\text{B14})$$

Equation (B14) gives a mean of

$$\langle x(t) \rangle = L \frac{t}{\langle\tau_1\rangle} + L \frac{1}{2} \frac{\sigma^2}{\langle\tau_1\rangle^2} - \langle\delta_x\rangle \quad (\text{B15})$$

and a variance of

$$\operatorname{var}[x(t)] = L^2 \frac{\sigma^2}{\langle\tau_1\rangle^2} \frac{t}{\langle\tau_1\rangle} + L^2 \frac{5}{4} \frac{\sigma^4}{\langle\tau_1\rangle^4} + \operatorname{var}(\delta_x). \quad (\text{B16})$$

As argued above, the statistics of δ_x are roughly independent of time on the timescale of period-to-period transport. In particular, we must have that $\langle\delta_x/t\rangle$ and $\operatorname{var}(\delta_x/t)$ converge to zero for large t .

At long times, we also find that Eq. (B14) converges to a normal distribution (Appendix B 4 a). In this limit, the constant terms in Eqs. (B15) and (B16) are negligible, and

$$\rho_t(x) \rightarrow \mathcal{N} \left(L \frac{t}{\langle\tau_1\rangle}, L^2 \frac{\sigma^2}{\langle\tau_1\rangle^2} \frac{t}{\langle\tau_1\rangle} \right). \quad (\text{B17})$$

In this form, it is clear that, on long timescales, distance is naturally counted in units of L , and time in units of $\langle\tau_1\rangle$ (see Lindner *et al.* [75] for related modeling).

Finally, it follows from Eqs. (B15) or (B17) that

$$\lim_{t \rightarrow \infty} \frac{\langle x(t) \rangle}{t} = \frac{L}{\langle\tau_1\rangle}. \quad (\text{B18})$$

Dividing both sides by Φ , we recover the desired result: Eq. (1) for the direct mobility is equivalent to Eq. (2) for the indirect mobility.

An additional result is that the quantity

$$D_{\text{eff}} = \frac{1}{2} \frac{\sigma^2}{\langle\tau_1\rangle^2} \frac{L^2}{\langle\tau_1\rangle} \quad (\text{B19})$$

behaves as an effective diffusion coefficient for the analyte. The ratio $\sigma/\langle\tau_1\rangle$ is almost the coefficient of variation of τ_1 . It differs because σ contains corrections due to the correlations between consecutive crossing times [Eq. (B8)]. Thus, we see that the model nicely reflects how correlations directly impact the dispersion of analytes as they travel through the system.

a. The limiting position distribution at long times is Gaussian

This section contains the derivation that the probability density function in Eq. (B12) converges to the probability density function of a normal distribution. Consider the shifted and scaled variable

$$q = \frac{k - \frac{t}{\mu}}{\frac{\sigma}{\mu} \sqrt{\frac{t}{\mu}}}, \quad (\text{B20})$$

which at large times will have a mean approaching zero and a variance approaching one. This has probability density function

$$\rho(q; t) = \frac{1 + \frac{1}{2} \frac{\sigma}{\sqrt{\mu t}} q}{\sqrt{2\pi} \left(1 + \frac{\sigma}{\sqrt{\mu t}} q\right)^3} \exp\left[-\frac{q^2}{2\left(1 + \frac{\sigma}{\sqrt{\mu t}} q\right)}\right]. \quad (\text{B21})$$

When considering values of q that are small compared to $\frac{\sqrt{\mu t}}{\sigma}$, the distribution will be very close to its limiting form of

$$\rho(q; t) \rightarrow \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{q^2}{2}\right) = \mathcal{N}(0, 1). \quad (\text{B22})$$

At any fixed t , no matter how large, the distribution of q will differ from this limiting form for sufficiently large q , i.e., in the distant tails of the distribution. However, given that the distribution is normalized, at very large t , the total probability assigned to these distant tails will be vanishingly small. The same derivation applies to the distribution in terms of $x + \delta_x$, rather than k [Eq. (B14)].

APPENDIX C: CONVERGENCE ANALYSIS

In this section, we will analyze and compare the numerical properties of the direct and indirect mobility formulations. Specifically, we will examine the computational cost of estimating each kind of mobility using molecular dynamics simulations of the analyte moving through the periodic geometry. Certain simplifying assumptions will be needed to advance the analysis. Most importantly, the analysis of the direct mobility will neglect the dynamics of δ_x (Appendix B 4), and the analysis of the indirect mobility will be based on the assumption that first-passage times are exponentially distributed at long times. The predicted scaling behaviors will be compared to numerical results on the example described in Sec. II B.

The underlying simulation implementation is assumed to be identical between the two cases, except for boundary and termination conditions. In particular, we will not consider the convergence of numerical error with respect to the discretization scheme. The error introduced by discretizing the equations of motion depends on the discretization scheme. Better schemes can be combined with either mobility formulation, and this consideration is essentially orthogonal to the comparison being made here. Of course, the discretization error may propagate differently in the simulations that would be used to calculate direct and indirect mobilities. Thus, in practice, some residual numerical error will always exist between the two.

1. Convergence of direct mobility

The direct mobility is typically estimated as

$$\hat{\mu}_{\text{direct}} = \frac{1}{\Phi t_{\text{direct}}} \left[\frac{1}{N_{\text{direct}}} \sum_{j=1}^{N_{\text{direct}}} x^{(j)} \right], \quad (\text{C1})$$

where N_{direct} molecules are simulated (independently) for a long period of time t_{direct} and the final states are used to estimate the direct mobility. In practice, t_{direct} is commonly chosen approximately as the time after which at least a certain

number of periods $k_{\text{direct}} \approx t_{\text{direct}} / \langle \tau_1 \rangle$ will have been traversed on average.

Equations (B15) and (B16) from Appendix B, allow us to predict the limiting behavior of the relative error between $\hat{\mu}_{\text{direct}}$ and the true direct mobility μ_{direct} . For large t_{direct} , each particle's position is identically and independently normally distributed with mean and variance given by Eqs. (B15) and (B16). The relative error is thus also normally distributed, with

$$\text{mean}\left(\frac{\hat{\mu}_{\text{direct}} - \mu_{\text{direct}}}{\mu_{\text{direct}}}\right) \approx \left(\frac{1}{2} \frac{\sigma^2}{\langle \tau_1 \rangle^2} - \frac{\langle \delta_x \rangle}{L}\right) \frac{1}{k_{\text{direct}}}, \quad (\text{C2})$$

$$\text{stderr}\left(\frac{\hat{\mu}_{\text{direct}} - \mu_{\text{direct}}}{\mu_{\text{direct}}}\right) \approx \frac{\sigma}{\langle \tau_1 \rangle} \frac{1}{\sqrt{N_{\text{direct}} k_{\text{direct}}}}, \quad (\text{C3})$$

where we have ignored terms of order $\mathcal{O}[1/(k_{\text{direct}} \sqrt{N_{\text{direct}}})]$ in the standard error. The mean relative error depends on the behavior of $\langle \delta_x \rangle$, which is outside the scope of the present study. In the following discussion, we will consider the simple case of $\langle \delta_x \rangle \approx 0$, which we expect to be reasonable for highly driven systems. As we will see in the numerical demonstrations (Sec. III C), $\langle \delta_x \rangle$ plays an important role in weakly driven systems.

The mean relative error [Eq. (C2)] indicates a bias due to the finite simulation time with which the direct mobility is being estimated. It cannot be reduced except by increasing k_{direct} , and it decays at a rate of $\mathcal{O}(1/k_{\text{direct}})$. Conversely, the standard relative error [Eq. (C3)] captures the intrinsic noise in the mobility estimator. This decays as $\mathcal{O}(1/\sqrt{k_{\text{direct}}})$, which is slower than the decay of the mean relative error. Thus, the direct mobility estimator will be statistically indistinguishable from an unbiased estimator for sufficiently long runtimes.

The limiting behavior of the direct mobility estimator is jointly affected by k_{direct} and N_{direct} . A reasonable choice (see, e.g., Mark and Baram [82]) for a single scalar error is the square root of the expected square of the relative error between $\hat{\mu}_{\text{direct}}$ and the true direct mobility μ_{direct} :

$$\epsilon_{\text{direct}}^2 := \mathbb{E}\left[\left(\frac{\hat{\mu}_{\text{direct}} - \mu_{\text{direct}}}{\mu_{\text{direct}}}\right)^2\right] \approx \text{bias}^2 + \text{stderr}^2, \quad (\text{C4})$$

where bias is the mean relative error given by Eq. (C2) and stderr is the standard relative error given by Eq. (C3).

2. Convergence of indirect mobility

Estimating the indirect mobility is tantamount to estimating the mean first-passage time of particles crossing a single period starting from the initial conditions $x = x_0$ and $\theta_0 \sim \pi(\theta)$. We will assume that the stationary distribution $\pi(\theta_0)$ is known and/or can be sampled efficiently. A careful cost analysis of this sampling process is relegated to future work.

First, let us consider estimating the indirect mobility using

$$\hat{\mu}_{\text{indirect}} = \frac{L}{\Phi} \left(\frac{1}{N_{\text{indirect}}} \sum_{j=1}^{N_{\text{indirect}}} \tau_1^{(j)} \right)^{-1}, \quad (\text{C5})$$

which is based on sampling N_{indirect} instances $\tau_1^{(j)}$ of the first-passage time. It may be difficult in general to ascertain whether this is an unbiased estimator of the indirect mobility. To make progress on this and other questions, we propose that it is reasonable to assume that the first-passage time across

one period is roughly exponentially distributed at long times. Thus, throughout the rest of this discussion, we will assume

$$\rho(\tau_1) \approx \frac{1}{\tau^*} \exp\left(-\frac{\tau_1}{\tau^*}\right) \quad (\text{C6})$$

at large τ_1 , where τ^* is some constant. Heuristically, this will be the case for systems in which the first-passage process converges to a steady-state behavior after initial transient behavior decays. It is consistent with the spirit of Kramers theory (Appendix A 2) since we are restricting our attention to the residual first-passage process on long timescales (compared to all other timescales of relaxation in the system). This can be justified more rigorously for many typical systems by considering the behavior of the eigenfunctions of the Smoluchowski equation in the presence of an absorber (e.g., as done by Grigoriev *et al.* [83] when studying the narrow escape problem); in this case, τ^* will be the first eigenvalue of the PDE. In particular, although the equivalence of indirect and direct mobilities was proven in Appendix B for any system, including those where transient dynamics are nonnegligible, our error analysis in this section will apply to convergence rates within the limit of long times after transient phenomena have abated.

In the case of an exponential distribution, it is known that the simple estimate used in Eq. (C5) is indeed a maximum likelihood estimator but is nonetheless biased. Specifically, the limiting relative error in the mobility estimator for exponentially distributed first-passage times is simply

$$\text{mean}\left(\frac{\hat{\mu}_{\text{indirect}} - \mu_{\text{indirect}}}{\mu_{\text{indirect}}}\right) \approx \frac{1}{N_{\text{indirect}}}. \quad (\text{C7})$$

In any case, since the N_{indirect} samples of $\tau_1^{(j)}$ are independent, we can estimate the standard error of the indirect mobility estimator in Eq. (C5). The error in the estimate of $\langle\tau_1\rangle$ will go as

$$\text{stderr}\left(\langle\tau_1\rangle - \frac{1}{N_{\text{indirect}}} \sum_{j=1}^{N_{\text{indirect}}} \tau_1^{(j)}\right) \approx \frac{\sigma_0}{\sqrt{N_{\text{indirect}}}}, \quad (\text{C8})$$

where we have introduced the notation $\sigma_0 := \text{stddev}(\tau_1)$ to indicate the standard deviation of τ_1 . Propagating the uncertainty therefore yields that

$$\text{stderr}\left(\frac{\hat{\mu}_{\text{indirect}} - \mu_{\text{indirect}}}{\mu_{\text{indirect}}}\right) = \frac{\sigma_0}{\langle\tau_1\rangle \sqrt{N_{\text{indirect}}}}. \quad (\text{C9})$$

Note that, as expected, the standard error of the indirect mobility estimator scales as $\mathcal{O}(1/\sqrt{N_{\text{indirect}}})$. This bias is negligible relative to the standard error for sufficiently large N_{indirect} . The total error will therefore also converge as

$$\epsilon_{\text{indirect}} \approx \frac{\sigma_0}{\langle\tau_1\rangle \sqrt{N_{\text{indirect}}}}. \quad (\text{C10})$$

In our experience, the coefficient of variation is of order one, and the bias is thus at most a 1% correction to Eq. (C10).

3. Comparing convergence: Entirely serial computation

First, let us analyze the runtimes of the direct and indirect mobility estimators in the case where all computations are performed in serial. Then the total runtime for the direct mobility

estimator will be proportional to

$$T_{\text{direct}} = t_{\text{direct}} N_{\text{direct}} \approx \langle\tau_1\rangle k_{\text{direct}} N_{\text{direct}}, \quad (\text{C11})$$

up to a constant factor based on the implementation of the simulations (i.e., the real time elapsed per unit of simulation time simulated). We will assume for the rest of this section that this proportionality factor is some constant and omit it from the discussion.

The error of the direct mobility estimator can be decreased by increasing either k_{direct} or N_{direct} . The runtime is linear in each of these, and the standard relative error depends equally on both quantities. However, the bias depends only on k_{direct} , and so the best choice in this circumstance is to fix $N_{\text{direct}} = 1$. Thus, using Eqs. (C11) and (C4), the runtime necessary to reach a small relative error ϵ scales as

$$\frac{T_{\text{direct}}}{\langle\tau_1\rangle} \approx \frac{1}{2} \frac{\sigma^2}{\langle\tau_1\rangle^2} \frac{1}{\epsilon^2} (1 + \sqrt{1 + \epsilon^2}) \rightarrow \frac{\sigma^2}{\langle\tau_1\rangle^2} \frac{1}{\epsilon^2}. \quad (\text{C12})$$

The term $1 + \sqrt{1 + \epsilon^2}$ is very nearly equal to two for reasonable values of ϵ (say, below 10%).

Similarly, the total runtime for the indirect mobility estimator will be approximately proportional to

$$T_{\text{indirect}} \approx \langle\tau_1\rangle N_{\text{indirect}} \quad (\text{C13})$$

for large N_{indirect} . Using Eqs. (C13) and (C10), we find that its runtime will therefore grow as

$$\frac{T_{\text{indirect}}}{\langle\tau_1\rangle} \approx \frac{\sigma_0^2}{\langle\tau_1\rangle^2} \frac{1}{\epsilon^2}. \quad (\text{C14})$$

Therefore, in the case of purely serial computing, the two formulations are nearly identical. Both have runtimes of order $\mathcal{O}(1/\epsilon^2)$. The convergence of the direct mobility estimator is influenced by crossing time correlations (via σ), whereas the indirect mobility estimator is not. The direct mobility estimator will exhibit more error in highly correlated systems than the indirect mobility estimator. However, correctly sampling the stationary distribution required for the indirect mobility estimator may become more difficult in such systems. In any case, neither algorithm is clearly advantageous when computations are done in serial.

4. Comparing convergence: Entirely parallel computation

Now, suppose instead that all sampling of trajectories will be computed entirely in parallel. Although the total number of floating-point operations will be the same as in the case of serial computation, in the case of parallel computation, it is often of more interest to consider the total elapsed time from the start of the algorithm to the termination of the last parallel thread of the computation.

For the direct mobility estimator, every parallel thread has the same fixed runtime (up to fluctuations in the computing speed). Thus, the total time elapsed will simply be proportional to the duration of each trajectory:

$$T_{\text{direct}} = t_{\text{direct}} \approx \langle\tau_1\rangle k_{\text{direct}}. \quad (\text{C15})$$

The standard relative error will be vanishingly small, so the error will be dominated by the bias. Thus, the runtime will

now converge as

$$\frac{T_{\text{direct}}}{\langle \tau_1 \rangle} \approx \frac{1}{2} \frac{\sigma^2}{\langle \tau_1 \rangle^2} \frac{1}{\epsilon}. \quad (\text{C16})$$

This is substantially better than in the case of purely serial computation [Eq. (C12)] and is now of order $\mathcal{O}(1/\epsilon)$.

Assessing the total time elapsed for the indirect mobility estimator is more complicated. Because the indirect mobility is independently sampling the first-passage time, each sample's runtime is a stochastic quantity. In the case of serial computation, the central limit theorem ensures that the estimate in Eq. (C13) will be fairly accurate for large N_{indirect} . However, if trajectories are computed entirely in parallel, then the elapsed time from start to finish will be dictated by the sample of the maximum first-passage time rather than the mean first passage time:

$$T_{\text{indirect}} \approx \max_{N_{\text{indirect}}} (\tau_1^{(j)}). \quad (\text{C17})$$

Naturally, the sample maximum will depend on the simulated ensemble size N_{indirect} .

The typical maximum first-passage time can be estimated if we again suppose that, at long times, the first-passage time is exponentially distributed. Setting the cumulative distribution of Eq. (C6) equal to $1 - (1/N_{\text{indirect}})$ yields the following estimate for the time at which the last particle will escape:

$$\frac{1}{N_{\text{indirect}}} \approx \exp\left(-\frac{T_{\text{indirect}}}{\tau^*}\right) \quad (\text{C18})$$

$$\Rightarrow T_{\text{indirect}} \approx \tau^* \ln(N_{\text{indirect}}). \quad (\text{C19})$$

A more rigorous justification for this runtime estimate can be obtained by considering the distribution of the maximum of an ensemble of i.i.d. exponentially distributed variables. By appealing to Poisson processes and harmonic numbers, one can recover the logarithmic scaling of T_{indirect} with N_{indirect} again. Incidentally, when the coefficient of variation of the first-passage time is greater than one, the argument to the logarithm in Eq. (C19) should, in fact, be the fraction of the population belonging to long tails of the distribution. However, because the dependence is logarithmic, the effect of this correction is small.

Note that this perspective implicitly assumes that all parallel computation hardware remains reserved for the respective calculations until all trajectories are completed. This is, in fact, true for the direct mobility estimator since all parallel computations will have the same runtime. Conversely, most samples generated towards the indirect mobility estimator will have runtimes much smaller than the maximum first-passage time. Thus, T_{indirect} is undoubtedly an overestimate of the computational cost in settings where parallel computing resources can be repurposed dynamically as soon as these samples terminate.

In any case, using Eq. (C19) for the runtime with Eq. (C10) for the error, we see that the indirect mobility estimator converges exponentially as

$$\epsilon \approx \frac{\sigma_0}{\langle \tau_1 \rangle} \exp\left(-\frac{1}{2} \frac{T_{\text{indirect}}}{\tau^*}\right). \quad (\text{C20})$$

Equivalently,

$$T_{\text{indirect}} \approx 2\tau^* \ln\left(\frac{\sigma_0}{\langle \tau_1 \rangle} \frac{1}{\epsilon}\right). \quad (\text{C21})$$

Thus, as ϵ becomes small and $1/\epsilon$ becomes larger, the required runtime grows only logarithmically. This will be exponentially faster than the convergence of the direct mobility estimator for small ϵ .

The results of this analysis have some caveats. The prefactor in Eq. (C16) is likely overestimated in general, because $\langle \delta_x \rangle$ has been ignored. Meanwhile, it is possible that $\tau^* \gg \langle \tau_1 \rangle$. These corrections will tend to improve the relative performance of the direct mobility estimator against the indirect mobility estimator, as is indeed observed in the numerical demonstrations of Sec. III C. Regardless, because the indirect mobility converges exponentially [Eq. (C20)], whereas the direct mobility converges as $T_{\text{direct}} \sim \mathcal{O}(1/\epsilon)$ [Eq. (C16)], these corrections are only important in comparing behavior at moderately large ϵ . When unlimited parallel computation is available, the indirect mobility formulation will always be much more efficient at sufficiently small ϵ .

5. Comparing convergence: Limited parallel computation

In practice, of course, unlimited parallelization is not feasible. Suppose that N_{para} samples can comfortably be simulated in parallel. For ensemble sizes larger than this, calculations must be broken into batches of N_{para} .

In this case, the direct mobility estimator's runtime will scale as

$$T_{\text{direct}} \approx \langle \tau_1 \rangle k_{\text{direct}} \left\lceil \frac{N_{\text{direct}}}{N_{\text{para}}} \right\rceil, \quad (\text{C22})$$

where $\lceil \cdot \rceil$ denotes the ceiling function. Because of the ceiling function, the runtime does not increase with the number of parallel trajectories until N_{direct} reaches an integral multiple of N_{para} . Thus, the best choice of N_{direct} is certainly at least N_{para} . However, going from $N_{\text{direct}} = N_{\text{para}}$ to $N_{\text{direct}} = 2N_{\text{para}}$ increases the runtime by a factor of two while leaving the bias [Eq. (C2)] unchanged. Increasing k_{direct} by a factor of two would have the same impact on runtime and standard relative error but would also decrease the bias. Mirroring the reasoning from Appendix C3, we thus find that the optimal choice is precisely $N_{\text{direct}} = N_{\text{para}}$.

Using Eq. (C4) with $N_{\text{direct}} = N_{\text{para}}$ and Eq. (C22) yields the runtime necessary to attain a target accuracy ϵ :

$$\frac{T_{\text{direct}}}{\langle \tau_1 \rangle} = \frac{1}{2} \frac{\sigma^2}{\langle \tau_1 \rangle^2} \frac{1}{\epsilon^2} \frac{1}{N_{\text{para}}} [1 + \sqrt{1 + (\epsilon N_{\text{para}})^2}]. \quad (\text{C23})$$

This is similar to the result for serial computation [Eq. (C12)] but differs in two places. First, the prefactor of $1/N_{\text{para}}$ corresponds to the acceleration of convergence by a factor of N_{para} in the small- ϵ limit. Here, the error is dominated by noise [Eq. (C3)] and $\epsilon \sim \mathcal{O}(1/\sqrt{T_{\text{direct}}})$. Increasing the number of independent samples is essentially as beneficial as increasing k_{direct} by the same amount.

However, the term $(\epsilon N_{\text{para}})^2$ inside the square root of Eq. (C23) corresponds to an acceleration at larger values of ϵ . Since the computational cost does not increase with N_{indirect} until $N_{\text{direct}} = N_{\text{para}}$, the convergence for $N_{\text{direct}} < N_{\text{para}}$

is essentially the same as in the case of unlimited parallel computation. Specifically, when $\epsilon N_{\text{para}} \gg 1$

$$1 + \sqrt{1 + (\epsilon N_{\text{para}})^2} \approx \epsilon N_{\text{para}}, \quad (\text{C24})$$

which implies that

$$\frac{T_{\text{direct}}}{\langle \tau_1 \rangle} \approx \frac{1}{2} \frac{\sigma^2}{\langle \tau_1 \rangle^2} \frac{1}{\epsilon}. \quad (\text{C25})$$

In other words, when $N_{\text{direct}} < N_{\text{para}}$, the error is dominated by bias [Eq. (C2)], and the convergence is of order $\epsilon \sim \mathcal{O}(1/T_{\text{direct}})$.

Consider now the indirect mobility estimator. Its runtime in the case of limited parallel computation depends on the manner in which it is implemented. We will consider two approaches. To enable these more complicated analyses, we will assume first-passage times are distributed exponentially as per Eq. (C19) with $\tau^* \approx \langle \tau_1 \rangle$. Accounting for deviations from a single exponential distribution makes these algorithms more challenging to analyze. Conservative approximations can be obtained by increasing the runtime estimates by $\tau^*/\langle \tau_1 \rangle$.

First, consider an algorithm for computing the indirect mobility estimator in which N_{indirect} trajectories are initiated at once and evolved in time at the same rate. The first N_{para} trajectories are incremented by one timestep, then the next N_{para} are incremented once, and so on until all trajectories have been incremented once. Early in the simulation, it will take $\lceil N_{\text{indirect}}/N_{\text{para}} \rceil$ passes to increment all trajectories by one timestep. As the simulation advances and some events terminate, fewer passes will be required to increment time. In this case, the runtime will scale as

$$\frac{T_{\text{indirect}}}{\langle \tau_1 \rangle} \approx \ln(N_{\text{last}}) \quad (\text{C26})$$

$$+ \sum_{k=1}^{M-1} \ln \left[\frac{N_{\text{indirect}} - (k-1)N_{\text{para}}}{N_{\text{indirect}} - kN_{\text{para}}} \right] [M - (k-1)] \quad (\text{C27})$$

$$= (4M - 2) \ln(M) - \ln(M!) + \ln(N_{\text{last}}), \quad (\text{C28})$$

$$\approx (3M - 2) \ln(M) + M + \frac{1}{2} \ln(2\pi M) + \ln(N_{\text{last}}), \quad (\text{C29})$$

where $M = \lceil N_{\text{indirect}}/N_{\text{para}} \rceil$ and $N_{\text{last}} = N_{\text{indirect}} - (M-1)N_{\text{para}}$. The factor of $\ln(N_{\text{last}})$ accounts for the maximum first-passage time in that final batch. The approximation in Eq. (C29) is based on Stirling's approximation, and reveals that this runtime is $\mathcal{O}[M \ln(M)]$.

Consider next an alternative implementation of the indirect mobility estimator with limited parallel computation. In this case, the simulation begins by initializing only N_{para} trajectories. A new trajectory is initiated on the same thread whenever a trajectory terminates until a total of N_{indirect} has been initiated. In this case, the runtime will scale as

$$\frac{T_{\text{indirect}}}{\langle \tau_1 \rangle} \approx \frac{N_{\text{indirect}} - N_{\text{min}}}{N_{\text{para}}} + \ln(N_{\text{min}}), \quad (\text{C30})$$

where $N_{\text{min}} = \min(N_{\text{indirect}}, N_{\text{para}})$. The first term estimates the time until a total of N_{indirect} trajectories have been initiated,

and the second term estimates the time required for the simulations to terminate thereafter. This algorithm's runtime is $\mathcal{O}(N_{\text{indirect}})$.

Surprisingly, this second implementation is consistently faster than the first one. There are two reasons for this. First, it ensures that no parallel computing threads are idle until the last batch of simulations when some idling is inevitable. More important, however, is that the second algorithm allows for a natural balancing of fast and slow events across different threads. Threads on which events terminate quickly will more quickly be reinitialized with new events. Conversely, in the first algorithm, all events are simulated independently so that the maximum first-passage time overall N_{indirect} events factors into the overall runtime. Worse, the speed at which these long trajectories are simulated is impaired by a factor of $\mathcal{O}(M)$ for most of the runtime. We will proceed with the analysis of the second, faster algorithm, but we include the analysis of the first version here as a warning to the reader.

Equation (C30) can be written more explicitly as

$$\frac{T_{\text{indirect}}}{\langle \tau_1 \rangle} \approx \begin{cases} \ln(N_{\text{indirect}}), & N_{\text{indirect}} \leq N_{\text{para}}, \\ \frac{N_{\text{indirect}}}{N_{\text{para}}} + \ln(N_{\text{para}}) - 1, & N_{\text{indirect}} > N_{\text{para}}. \end{cases} \quad (\text{C31})$$

Reversing Eq. (C10), we find that the number of samples necessary for the indirect mobility estimator to achieve a target relative error ϵ is

$$N_{\text{indirect}} = \frac{\sigma_0^2}{\langle \tau_1 \rangle^2} \frac{1}{\epsilon^2}. \quad (\text{C32})$$

Substituting Eq. (C32) into Eq. (C31) yields the runtime required to achieve the target accuracy. When $N_{\text{indirect}} \leq N_{\text{para}}$,

$$\frac{T_{\text{indirect}}}{\langle \tau_1 \rangle} \approx 2 \ln \left(\frac{\sigma_0}{\langle \tau_1 \rangle} \frac{1}{\epsilon} \right). \quad (\text{C33})$$

This is equivalent to Eq. (C20). In this regime, the error behaves as if there were unlimited parallelism and so decreases exponentially with runtime.

Conversely, when $N_{\text{indirect}} > N_{\text{para}}$,

$$\frac{T_{\text{indirect}}}{\langle \tau_1 \rangle} \approx \frac{\sigma_0^2}{\langle \tau_1 \rangle^2} \frac{1}{\epsilon^2} \frac{1}{N_{\text{para}}} + \ln(N_{\text{para}}) - 1. \quad (\text{C34})$$

This is very similar to the result for entirely serial computation [Eq. (C14)], but accelerated by a factor of N_{para} . The $\ln(N_{\text{para}})$ term arises here because the large parallel batches are more vulnerable to rare long-duration events. However, this term is negligible in the limit of $N_{\text{indirect}} \gg N_{\text{para}}$ and the error scales as $\mathcal{O}(1/\sqrt{T_{\text{indirect}}})$, as expected. In fact, on the limit of small target error and/or small N_{para} , the two mobility formulations are once again essentially equivalent up to a factor of $\psi = \sigma/\sigma_0$.

Figure 7 in the main body of the paper summarizes the theoretically predicted ratio of runtimes for the direct and indirect mobilities; ψ was factored out from the direct mobility runtime. Figure 7(a) shows the predicted ratio as a function of the final relative error ϵ . The colors blue, orange, green correspond to $N_{\text{para}} = 10^5, 10^4, 10^3$, respectively. A typical modern consumer-grade GPU can effectively deliver tens of thousands of parallel threads, justifying $N_{\text{para}} = 10^4$ for single-body molecule simulations. Parallelizing across ten such GPUs is reasonably economical in many cases, motivating the case $N_{\text{para}} = 10^5$. Conversely, many-body molecules will reduce

the number of independent simulations that can be conducted in parallel, which motivated the choice of $N_{\text{para}} = 10^3$. The different line styles correspond to coefficients of variation equal to 3 (solid), 1 (dashed), and 0.5 (dotted), which are values encountered in the example system from Sec. II B [23]. The lines are truncated at the values of ϵ such that either k_{direct} or N_{indirect} would be required to equal 10 or less; the various modeling assumptions certainly do not apply in that regime.

Altogether, the theoretical analysis predicts that the indirect mobility estimator will converge up to 2–70 times faster than the direct mobility estimator under these circumstances. The ϵ of maximum relative advantage is in the range of 0.1–1%. This is often a perfectly acceptable error threshold for assisting with the research and design of periodic MNFDs, as modeling errors are often larger than this. As noted, at very small target relative errors, the estimators are essentially equivalent (i.e., the ratio converges to 1).

Figures 7(b) and 7(c) highlight the expected transitions of the direct and indirect mobility estimators, respectively, from parallel-like scaling to serial-like scaling. For the direct mobility estimator, this occurs at $\epsilon \approx 1/N_{\text{para}}$, where for the indirect mobility estimator it occurs near $\epsilon \approx (\sigma_0/\langle\tau_1\rangle)/\sqrt{N_{\text{para}}}$. The ϵ value of maximum relative advantage for the indirect mobility estimator is expected to occur somewhere between these two points.

As noted at the end of Appendix C 3, the net effect of the correlation factor ψ does not clearly favour either algorithm. Although it directly amplifies the predicted runtime of the direct mobility estimator, it likely also increases the cost of sampling the stationary distribution. In any case, it does not appear that large correlation functions are likely to occur in most applications, so this effect is likely to be modest. If there are applications of interest where ψ is found to be large, then this aspect of the algorithm should be investigated in more detail.

-
- [1] K. D. Dorfman, DNA electrophoresis in microfabricated devices, *Rev. Mod. Phys.* **82**, 2903 (2010).
 - [2] S. L. Levy and H. G. Craighead, DNA manipulation, sorting, and mapping in nanofluidic systems, *Chem. Soc. Rev.* **39**, 1133 (2010).
 - [3] K. D. Dorfman, S. B. King, D. W. Olson, J. D. P. Thomas, and D. R. Tree, Beyond gel electrophoresis: Microfluidic separations, fluorescence burst analysis, and DNA stretching, *Chem. Rev.* **113**, 2584 (2013).
 - [4] M. Sonker, D. Kim, A. Egatz-Gomez, and A. Ros, Separation phenomena in tailored micro- and nanofluidic environments, *Annu. Rev. Anal. Chem.* **12**, 475 (2019).
 - [5] W. D. Volkmuth and R. H. Austin, DNA electrophoresis in microlithographic arrays, *Nature* **358**, 600 (1992).
 - [6] K. D. Dorfman, DNA electrophoresis in microfluidic post arrays under moderate electric fields, *Phys. Rev. E* **73**, 061922 (2006).
 - [7] J. Cho, S. Kumar, and K. D. Dorfman, Electrophoretic collision of a DNA molecule with a small elliptical obstacle, *Electrophoresis* **31**, 860 (2010).
 - [8] P. K. Ghosh, P. Hänggi, F. Marchesoni, S. Martens, F. Nori, L. Schimansky-Geier, and G. Schmid, Driven Brownian transport through arrays of symmetric obstacles, *Phys. Rev. E* **85**, 011101 (2012).
 - [9] Z. Benková, L. Rišpanová, and P. Cifra, Effect of chain stiffness for semiflexible macromolecules in array of cylindrical nanoposts, *J. Chem. Phys.* **147**, 134907 (2017).
 - [10] M. C. VandeSande, D. J. Pasut, and H. W. de Haan, Sorting polymers by size via an array of viscous posts, *Electrophoresis* **38**, 2488 (2017).
 - [11] Q. Wu, N. Kaji, T. Yasui, S. Rahong, T. Yanagida, M. Kanai, K. Nagashima, M. Tokeshi, T. Kawai, and Y. Baba, A millisecond micro-RNA separation technique by a hybrid structure of nanopillars and nanoslits, *Sci. Rep.* **7**, 43877 (2017).
 - [12] T. Ajiri, T. Yasui, M. Maeki, A. Ishida, H. Tani, J. Nishii, Y. Baba, and M. Tokeshi, Silica nanopillar arrays for monitoring diffraction-based label-free biomolecule separation, *ACS Appl. Nano Mater.* **3**, 8810 (2020).
 - [13] B. Chakrabarti, C. Gaillard, and D. Saintillan, Trapping, gliding, vaulting: transport of semiflexible polymers in periodic post arrays, *Soft Matter* **16**, 5534 (2020).
 - [14] T. A. J. Duke and R. H. Austin, Microfabricated Sieve for the Continuous Sorting of Macromolecules, *Phys. Rev. Lett.* **80**, 1552 (1998).
 - [15] D. Ertaş, Lateral Separation of Macromolecules and Polyelectrolytes in Microlithographic Arrays, *Phys. Rev. Lett.* **80**, 1548 (1998).
 - [16] C.-F. Chou, O. Bakajin, S. W. P. Turner, T. A. J. Duke, S. S. Chan, E. C. Cox, H. G. Craighead, and R. H. Austin, Sorting by diffusion: An asymmetric obstacle course for continuous molecular separation, *Proc. Natl. Acad. Sci. USA* **96**, 13762 (1999).
 - [17] A. van Oudenaarden and S. G. Boxer, Brownian ratchets: Molecular separations in lipid bilayers supported on patterned arrays, *Science* **285**, 1046 (1999).
 - [18] J. Han and H. G. Craighead, Entropic trapping and sieving of long DNA molecules in a nanofluidic channel, *J. Vac. Sci. Technol. A* **17**, 2142 (1999).
 - [19] J. Han, S. W. Turner, and H. G. Craighead, Entropic Trapping and Escape of long DNA Molecules at Submicron size Constriction, *Phys. Rev. Lett.* **83**, 1688 (1999).
 - [20] J. Han and H. G. Craighead, Separation of long DNA molecules in a microfabricated entropic trap array, *Science* **288**, 1026 (2000).
 - [21] J. Han and H. G. Craighead, Characterization and optimization of an entropic trap for DNA separation, *Anal. Chem.* **74**, 394 (2002).
 - [22] N. Laachi, C. Declet, C. Matson, and K. D. Dorfman, Nonequilibrium Transport of Rigid Macromolecules in Periodically Constricted Geometries, *Phys. Rev. Lett.* **98**, 098106 (2007).
 - [23] K.-L. Cheng, Y.-J. Sheng, S. Jiang, and H.-K. Tsao, Electrophoretic size separation of particles in a periodically constricted microchannel, *J. Chem. Phys.* **128**, 101101 (2008).
 - [24] E. A. Strychalski, H. W. Lau, and L. A. Archer, Nonequilibrium separation of short DNA using nanoslit arrays, *J. Appl. Phys.* **106**, 024915 (2009).

- [25] Z. R. Li, G. R. Liu, N. G. Hadjiconstantinou, J. Han, J.-S. Wang, and Y. Z. Chen, Dispersive transport of biomolecules in periodic energy landscapes with application to nanofilter sieving arrays, *Electrophoresis* **32**, 506 (2011).
- [26] M. B. Mikkelsen, W. Reisner, H. Flyvbjerg, and, A. Kristensen, Pressure-driven DNA in nanogroove arrays: complex dynamics leads to length-and topology-dependent separation, *Nano Lett.* **11**, 1598 (2011).
- [27] S. T. T. Ollila, C. Denniston, M. Karttunen, and T. Ala-Nissila, Biopolymer Filtration in Corrugated Nanochannels, *Phys. Rev. Lett.* **112**, 118301 (2014).
- [28] L. Wu and S. Levy, Fluctuations of DNA mobility in nanofluidic entropic traps, *Biomicrofluidics* **8**, 044103 (2014).
- [29] M. Alishahi, R. Kamali, and O. Abouali, Numerical investigation of molecular nano-array in potential-energy profile for a single dsDNA, *Eur. Phys. J. E* **39**, 50 (2016).
- [30] S. H. Ko, D. Chandra, W. Ouyang, T. Kwon, P. Karande, and J. Han, Nanofluidic device for continuous multiparameter quality assurance of biologics, *Nat. Nanotechnol.* **12**, 804 (2017).
- [31] Z. Cao and L. Yobas, Fast DNA sieving through submicrometer cylindrical glass capillary matrix, *Anal. Chem.* **86**, 737 (2014).
- [32] Z. Cao and L. Yobas, Gel-free electrophoresis of DNA and proteins on chips featuring a 70 nm capillary-well motif, *ACS Nano* **9**, 427 (2015).
- [33] D. Nykypanchuk, H. H. Strey, and D. A. Hoagland, Brownian motion of DNA confined within a two-dimensional array, *Science* **297**, 987 (2002).
- [34] A. M. Berezhkovskii, V. Y. Zitserman, and S. Y. Shvartsman, Diffusivity in periodic arrays of spherical cavities, *J. Chem. Phys.* **118**, 7146 (2003).
- [35] A. M. Berezhkovskii, V. Y. Zitserman, and S. Y. Shvartsman, Effective diffusivity in periodic porous materials, *J. Chem. Phys.* **119**, 6991 (2003).
- [36] E. J. Saltzman and M. Muthukumar, Conformation and dynamics of model polymer in connected chamber-pore system, *J. Chem. Phys.* **131**, 214903 (2009).
- [37] K. Nagarajan and S. B. Chen, Driven transport of dilute polymer solutions through porous media comprising interconnected cavities, *Colloids Interfaces* **5**, 22 (2021).
- [38] T. Ikonen, Driven polymer transport through a periodically patterned channel, *J. Chem. Phys.* **140**, 234906 (2014).
- [39] Y.-C. Chen, Y.-L. Zhou, and C. Wang, Monte Carlo simulation on the diffusion of polymer in narrow periodical channels, *Int. J. Mod. Phys. B* **31**, 1750144 (2017).
- [40] L. B. Weiss, A. Nikoubashman, and C. N. Likos, Topology-sensitive microfluidic filter for polymers of varying stiffness, *ACS Macro Lett.* **6**, 1426 (2017).
- [41] C. Wang, Y.-L. Zhou, L.-Z. Sun, Y.-C. Chen, and M.-B. Luo, Simulation study on the migration of diblock copolymers in periodically patterned slits, *J. Chem. Phys.* **150**, 164904 (2019).
- [42] R. Riehn, R. H. Austin, and J. C. Sturm, A nanofluidic railroad switch for DNA, *Nano Lett.* **6**, 1973 (2006).
- [43] D. Kim, C. Bowman, J. T. Del Bonis-O'Donnell, A. Matzavinos, and D. Stein, Giant Acceleration of DNA Diffusion in an Array of Entropic Barriers, *Phys. Rev. Lett.* **118**, 048002 (2017).
- [44] M. Magill, E. Waller, and H. W. de Haan, A sequential nanopore-channel device for polymer separation, *J. Chem. Phys.* **149**, 174903 (2018).
- [45] Z. He and R. Zhou, Exploring an in-plane graphene and hexagonal boron nitride array for separation of single nucleotides, *ACS Nano* **15**, 11704 (2021).
- [46] E. Chmela, R. Tijssen, M. T. Blom, H. J. G. E. Gardeniers, and A. van den Berg, A chip system for size separation of macromolecules and particles by hydrodynamic chromatography, *Anal. Chem.* **74**, 3470 (2002).
- [47] D. Huh, J. H. Bahng, Y. Ling, H.-H. Wei, O. D. Kripfgans, J. B. Fowlkes, J. B. Grothberg, and S. Takayama, Gravity-driven microfluidic particle sorting device with hydrodynamic separation amplification, *Anal. Chem.* **79**, 1369 (2007).
- [48] L.-J. C. Jellema, A. P. Markesteijn, J. Westerweel, and E. Verpoorte, Tunable hydrodynamic chromatography of microparticles localized in short microchannels, *Anal. Chem.* **82**, 4027 (2010).
- [49] V. Biagioni, A. L. Sow, A. Adrover, and S. Cerbelli, Brownian sieving effect for boosting the performance of microcapillary hydrodynamic chromatography. Proof of concept, *Anal. Chem.* **93**, 6808 (2021).
- [50] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: Fifty years after Kramers, *Rev. Mod. Phys.* **62**, 251 (1990).
- [51] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer-Verlag, Berlin, 1992).
- [52] A. M. Nagel, M. Magill, and H. W. de Haan, Studying first-passage problems using neural networks: A case study in the slit-well microfluidic device, *Phys. Rev. E* **106**, 025311 (2022).
- [53] How many concurrent threads are running on my GeForce GTX 1080 Ti? NVIDIA Developer Forum (2022), retrieved from <https://forums.developer.nvidia.com/t/how-many-concurrent-threads-are-running-on-my-geforce-gtx-1080-ti/56708/3>
- [54] M. H. Jacobs, Diffusion processes, in *Diffusion Processes* (Springer, Berlin, 1967), pp. 1–145.
- [55] R. Zwanzig, Diffusion past an entropy barrier, *J. Phys. Chem.* **96**, 3926 (1992).
- [56] D. Reguera and J. M. Rubi, Kinetic equations for diffusion in the presence of entropic barriers, *Phys. Rev. E* **64**, 061106 (2001).
- [57] P. Kalinay and J. K. Percus, Corrections to the Fick-Jacobs equation, *Phys. Rev. E* **74**, 041203 (2006).
- [58] D. Reguera, G. Schmid, P. S. Burada, J. M. Rubi, P. Reimann, and P. Hänggi, Entropic Transport: Kinetics, Scaling, and Control Mechanisms, *Phys. Rev. Lett.* **96**, 130603 (2006).
- [59] N. Laachi, M. Kenward, E. Yariv, and K. D. Dorfman, Force-driven Transport through Periodic entropy Barriers, *Europhys. Lett.* **80**, 50009 (2007).
- [60] M.-V. Vazquez, A. M. Berezhkovskii, and L. Dagdug, Diffusion in linear porous media with periodic entropy barriers: A tube formed by contacting spheres, *J. Chem. Phys.* **129**, 046101 (2008).
- [61] G. Forte, F. Cecconi, and A. Vulpiani, Transport and fluctuation-dissipation relations in asymptotic and preasymptotic diffusion across channels with variable section, *Phys. Rev. E* **90**, 062110 (2014).
- [62] D. Reguera and J. M. Rubi, Engineering tube shapes to control confined transport, *Eur. Phys. J.: Spec. Top.* **223**, 3079 (2014).
- [63] J. Miguel Rubi, Entropic diffusion in confined soft-matter and biological systems, *Europhys. Lett.* **127**, 10001 (2019).
- [64] I. Pompa-García and L. Dagdug, Two-dimensional diffusion biased by a transverse gravitational force in an asymmetric

- channel: Reduction to an effective one-dimensional description, *Phys. Rev. E* **104**, 044118 (2021).
- [65] P. S. Burada, G. Schmid, D. Reguera, J. M. Rubi, and P. Hänggi, Biased diffusion in confined media: test of the Fick-Jacobs approximation and validity criteria, *Phys. Rev. E* **75**, 051111 (2007).
- [66] Y. A. Makhnovskii, V. Y. Zitserman, and A. M. Berezhkovskii, Diffusion in quasi-one-dimensional structures with a periodic sharp narrowing of the cross section, *Russ. J. Phys. Chem. B* **3**, 313 (2009).
- [67] A. M. Berezhkovskii and L. Dagdug, Analytical treatment of biased diffusion in tubes with periodic dead ends, *J. Chem. Phys.* **134**, 124109 (2011).
- [68] Y. A. Makhnovskii, A. M. Berezhkovskii, L. V. Bogachev, and V. Y. Zitserman, Driven diffusion in a periodically compartmentalized tube: Homogeneity versus intermittency of particle motion, *J. Phys. Chem. B* **115**, 3992 (2011).
- [69] A. M. Berezhkovskii and L. Dagdug, Biased diffusion in periodic potentials: Three types of force dependence of effective diffusivity and generalized Lifson-Jackson formula, *J. Chem. Phys.* **151**, 131102 (2019).
- [70] N. Khatri and P. S. Burada, Diffusion of interacting particles in a channel with reflection boundary conditions, *J. Chem. Phys.* **151**, 094103 (2019).
- [71] H.-W. Hu, L. Du, L.-H. Qu, Z.-L. Cao, Z.-C. Deng, and Y.-C. Lai, Current reversal and particle separation in Brownian transport, *Phys. Rev. Res.* **3**, 033162 (2021).
- [72] L. Dagdug, A. M. Berezhkovskii, V. Y. Zitserman, and S. M. Bezrukov, Effective diffusivity of a Brownian particle in a two-dimensional periodic channel of abruptly alternating width, *Phys. Rev. E* **103**, 062106 (2021).
- [73] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, J. M. Rubi, and A. Pérez-Madrid, Giant Acceleration of free Diffusion by use of Tilted Periodic Potentials, *Phys. Rev. Lett.* **87**, 010602 (2001).
- [74] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, J. M. Rubi, and A. Pérez-Madrid, Diffusion in tilted periodic potentials: Enhancement, universality, and scaling, *Phys. Rev. E* **65**, 031104 (2002).
- [75] B. Lindner, M. Kostur, and L. Schimansky-Geier, Optimal diffusive transport in a tilted periodic potential, *Fluctuation and Noise Letters* (World Scientific, Singapore, 2001).
- [76] P. I. Kuznetsov, R. L. Stratonovich, and V. I. Tikhonov, *Non-linear Transformations of Stochastic Processes* (Pergamon, Oxford, UK, 1965).
- [77] S. Lifson and J. L. Jackson, On the self-diffusion of ions in a polyelectrolyte solution, *J. Chem. Phys.* **36**, 2410 (1962).
- [78] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, UK, 2001).
- [79] S. Redner *et al.*, *A Guide to First-passage Processes* (Cambridge University Press, Cambridge, UK, 2001).
- [80] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953), Vol. 7.
- [81] T. E. Harris, The existence of stationary measures for certain Markov processes, *Matematika* **4**, 131 (1960); *Proc. 3rd Berkeley Sympos. Math. Statist. Probability*, **2**, 113 (1956).
- [82] Z. Mark and Y. Baram, The bias-variance dilemma of the Monte Carlo method, in *Proceedings of the International Conference on Artificial Neural Networks* (Springer, Berlin, 2001), pp. 141–147.
- [83] I. V. Grigoriev, Y. A. Makhnovskii, A. M. Berezhkovskii, and V. Y. Zitserman, Kinetics of escape through a small hole, *J. Chem. Phys.* **116**, 9574 (2002).