Evaluation of the Kirkwood approximation for the diffusivity of channel-confined DNA chains in the de Gennes regime

Aashish Jain and Kevin D. Dorfman[a)]
Department of Chemical Engineering and Material Science, University of Minnesota—Twin Cities, 421 Washington Ave. SE, Minneapolis, Minnesota 55455, USA

(Received 19 March 2015; accepted 30 March 2015; published online 7 April 2015)

We use Brownian dynamics with hydrodynamic interactions to calculate both the Kirkwood (short-time) diffusivity and the long-time diffusivity of DNA chains from free solution down to channel confinement in the de Gennes regime. The Kirkwood diffusivity in confinement is always higher than the diffusivity obtained from the mean-squared displacement of the center-of-mass, as is the case in free solution. Moreover, the divergence of the local diffusion tensor, which is non-zero in confinement, makes a negligible contribution to the latter diffusivity in confinement. The maximum error in the Kirkwood approximation in our simulations is about 2% for experimentally relevant simulation times. The error decreases with increasing confinement, consistent with arguments from blob theory and the molecular-weight dependence of the error in free solution. In light of the typical experimental errors in measuring the properties of channel-confined DNA, our results suggest that the Kirkwood approximation is sufficiently accurate to model experimental data. © *2015 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4917269]

I. INTRODUCTION

The dynamics of DNA in confinement play an important role in biology and biotechnology, ranging from the processes governing ejection of DNA from a viral capsid[1] to electrophoretic separations of DNA.[2] We are particularly interested in the dynamics of DNA molecules confined in nanochannels,[3] which form the basis for a new method of genome mapping.[4] In this method, DNA molecules containing sequence-specific fluorescent probes are stretched in an array of nanochannels, and the distance between probes is measured by fluorescence microscopy. The dynamic fluctuations of these probes under thermal energy sets the upper bound on the accuracy of the genomic measurements. As a result, the further advancement of genome mapping technology would benefit greatly from models that can predict these fluctuations. In particular, we would like to have a model that can accurately predict the center-of-mass diffusivity of a DNA chain in a nanochannel at a reasonable computational cost, which can then be used to model the relaxation time of the probe-laden DNA.[5–7]

Quantitative modeling of the dynamics of channel-confined DNA is challenging due to the need to account for the hydrodynamic interactions (HI) between the DNA segments and the channel walls.[8] While hydrodynamic screening by the walls greatly simplifies scaling theories[9] for the dynamics by removing hydrodynamic interactions between blobs, obtaining the prefactors for the scaling laws is complicated by the absence of a Green's function for the Stokes equation in confinement[8] and the cost of incorporating hydrodynamic interactions in general. Consequently, dynamic simulations of DNA in a nanochannel that incorporate hydrodynamic interactions[10–12] are restricted to relatively short chains or rather coarse models. The results thus obtained are informative, but the computational cost required to obtain them makes these methods prohibitive for studying DNA diffusion with the sub-persistence length resolution and long chains[13] required to address genome mapping technologies.

[a)]Electronic mail: dorfman@umn.edu

A possible alternative to dynamic simulations is the Kirkwood-Riseman[14,15] approach, commonly referred to as the Kirkwood approximation, which leads to a computationally efficient simulation method at the expense of an error in the computation of the diffusion coefficient. The so-called Kirkwood diffusivity is obtained by computing the double-sum

$$D^{(K)} = \frac{1}{3N_b^2}\sum_i \sum_j \mathrm{Tr}\langle \mathbf{D}_{ij}\rangle \tag{1}$$

over the N_b beads used to represent the chain. In Eq. (1), $\mathrm{Tr}\langle \mathbf{D}_{ij}\rangle$ denotes an ensemble average of the trace of the 3×3 block matrix that describes the hydrodynamic interaction (HI) between segments i and j in the $3N_b \times 3N_b$ diffusion tensor.[10] The ensemble averaging operator in Eq. (1) explains the computational speed of the Kirkwood approximation; the chain configurations used to compute $\mathrm{Tr}\langle \mathbf{D}_{ij}\rangle$ only need to be drawn from a suitable equilibrium ensemble. This ensemble can be generated, of course, from a dynamic simulation that incorporates hydrodynamic interactions.[10] However, the ensemble of chain configurations can be generated more quickly by a dynamic simulation without hydrodynamic interactions or, even better, by a suitable Monte Carlo method that rapidly explores the configurational phase space.[16] Once an equilibrium ensemble of configurations is generated, the diffusivity tensor is computed for each configuration and the results are averaged with respect to the weight of each configuration in the ensemble. This decoupling of the hydrodynamics and the chain configurations makes the Kirkwood approximation especially attractive for computing the diffusivity of DNA confined in nanochannels[13] and nanoslits[17] where hydrodynamic interactions with the walls greatly slow down the sampling of the configurational space in a dynamic simulation.

The improvement in computational speed in the Kirkwood approximation comes at the expense of accuracy. The Kirkwood approximation is essentially a hydrodynamic mean-field approximation, where a chain segment feels the average HI at every point in time. Accordingly, the approximation neglects the dynamic correlations that exist between the intramolecular hydrodynamic interactions at different points in time. Because these correlations are also small at short times, $D^{(K)}$ in Eq. (1) can be thought of as a short-time diffusivity. Fixman[18] showed that the Kirkwood approximation is always an overestimate of the long-time diffusivity computed from the Einstein relationship,

$$D_L = \lim_{t\to\infty} \frac{1}{6t}\langle \mathbf{R}_{\mathrm{cm}}^2\rangle, \tag{2}$$

for the displacement of the center-of-mass of the chain from its initial position at $\mathbf{R}_{\mathrm{cm}} = \mathbf{0}$. In Eq. (2) and hereafter, we interpret the square of a vector as $\mathbf{R}^2 = \mathbf{R}\cdot\mathbf{R}$, not as a dyadic product $\mathbf{RR}$. The difference between $D^{(K)}$ and D_L has been addressed for flexible and semiflexible polymer chains in free solution,[18–26] with errors in the range of 1% to 25% for different approaches and different polymer models. For example, the error increases by increasing the chain size,[24] or by increasing the flexibility of the chain[26] or by reducing the solvent quality.[23,25]

Our goal here is to evaluate the ability of the Kirkwood approximation to model experimental data for DNA confinement, for example, the relaxation time data from Reisner *et al.*[27] The total duration of these experiments is long compared to the autocorrelation time of the mean span of the chain, but is rarely longer than the time τ for center-of-mass diffusion over the size of the chain. As we will see, there is a fast transient in $\langle \mathbf{R}_{\mathrm{cm}}^2\rangle$ that captures much of the difference between the Kirkwood diffusion coefficient and the long-time diffusion coefficient. However, there may be an additional, very slow transient before the diffusivity finally reaches the asymptotic limit in Eq. (2). For example, simulations in free solution[24] suggested that very long simulations, at least 100τ in duration, are required to reach the asymptotic limit. As we will show later, such long simulations are infeasible for channel-confined DNA due to the cost of incorporating bead-wall hydrodynamic interactions. Fortunately, such long simulations are also irrelevant for the description of experimental data, since the experiments themselves only correspond to a few τ. So long as the simulations are sufficiently long to capture

the primary transient in $D^{(K)} - D(t)$, which will be the case here, any additional slow decay of the diffusivity towards its asymptotic value D_L will prove to be irrelevant to the analysis of experimental data.

The accuracy of the Kirkwood approximation in confinement has not been addressed completely to date. On the simulation side, Jendrejack *et al.*[10] demonstrated qualitative agreement between the Kirkwood approximation and the displacement over the center-of-mass trajectory for a relatively short simulation. This result is promising, but it leaves open questions surrounding the quantitative accuracy of the Kirkwood approximation in confinement, in particular, the effects of channel size, molecular weight, and the duration of the simulation. On the experimental side, there is one piece of indirect evidence[7] suggesting that the Kirkwood approximation is reasonable in confinement, namely, the agreement between experimental data for the longest relaxation time of channel-confined DNA[27] and a dumbbell model[7] whose friction coefficient was parameterized using the Kirkwood approximation.[13] However, the latter study[7] is not a direct test of the accuracy of the Kirkwood approximation since the comparison of the relaxation-time model with experiment conflates the issues surrounding the accuracy of the Kirkwood approximation in confinement with uncertainties arising from sampling of the chain configurations and the experimental data themselves.

In the present contribution, we provide a direct test of the accuracy of the Kirkwood approximation for channel-confined DNA. We limit ourselves to confinement no stronger than that in the de Gennes regime,[28] corresponding to channel sizes H satisfying $R_g \gtrsim H \gtrsim l_p^2/w$, where R_g is the radius of gyration of the DNA in free solution, l_p is the persistence length, and w is the effective width.[29,30] We simulate an experimentally validated model for DNA in such weak confinement[31] using Brownian dynamics (BD) with hydrodynamic interactions (HIs),[10] which provides temporal data for the chain configuration and the forces acting on the DNA segments. We use the correlations in the velocity perturbation and stochastic fluctuations data obtained from the configuration and force data to compute the long-time diffusivity D_L through a modification of the method of Liu and Dünweg,[24] while the configurational data allow us to compute the Kirkwood diffusivity. We are thus able to compare $D^{(K)}$ and D_L using the same ensemble of configurations, providing the desired direct test of the Kirkwood approximation. Our results indicate that the Kirkwood approximation improves as confinement increases, which we rationalize through the hydrodynamic screening between blobs in the de Gennes regime.

II. DNA MODEL AND SIMULATION ALGORITHM

A. DNA model

Our simulations use the bead-spring model of Jendrejack *et al.*,[10] which was parameterized for DNA in free solution[32] and provides good agreement with experiments on DNA confinement in the de Gennes regime in slits.[31] The model represents the DNA as N_b beads connected through $N_s = N_b - 1$ entropic springs. Each spring represents $N_{k,s} = N_k/N_s$ Kuhn lengths, where $N_k = L/b_k$ is the number of Kuhn segments of length b_k in a chain of contour length L. We follow the discretization of Jendrejack *et al.*[10] and use $N_{k,s} = 19.8$. This is a rather coarse model to use in confinement, which limits its applicability to the de Gennes regime but allows dynamic simulations of experimentally relevant molecular weights in such channels.

The semiflexibility of DNA appears in this model through Marko-Siggia spring force $\mathbf{F}_{ij}^{\mathrm{S}}$ between bonded beads,[33,34]

$$\mathbf{F}_{ij}^{\mathrm{S}} = \frac{k_B T}{2b_k}\left[\left(1 - \frac{R_{ij}}{q_0}\right)^{-2} - 1 + 4\frac{R_{ij}}{q_0}\right]\frac{\mathbf{R_{ij}}}{R_{ij}}, \tag{3}$$

where k_B is Boltzmann's constant, T is the absolute temperature, $q_0 = N_{k,s} b_k$ is the maximum spring length, $\mathbf{R}_{ij} = \mathbf{R}_j - \mathbf{R}_i$ is the connector vector between beads i and j, and R_{ij} is the magnitude of the connector vector. Here, $\mathbf{R}_i$ and $\mathbf{R}_j$ denote the absolute position vectors of beads i

and j, respectively. Bead-bead excluded volume interactions are modeled with the narrow Gaussian potential[10]

$$\mathbf{F}_{ij}^{\mathrm{EV}} = \pi \phi k_B T N_{k,s}^2 \left(\frac{9}{2b_k^2 N_{k,s} \pi}\right)^{5/2} \exp\left(\frac{-9R_{ij}^2}{2N_{k,s}b_k^2}\right) \mathbf{R_{ij}}, \tag{4}$$

where ϕ is the excluded volume parameter. Bead-wall interactions are governed by a different potential[10]

$$U_i^w = \begin{cases} \dfrac{25k_BT}{3b_k\delta_w^2}(h-\delta_w)^3 & \text{for } h < \delta_w = \dfrac{b_k\sqrt{N_{k,s}}}{2}, \\ 0 & \text{for } h \geq \delta_w, \end{cases} \tag{5}$$

where h is the distance of bead i from the wall into the fluid in the wall-normal direction.

In addition to the latter potentials, which govern the thermodynamic properties of the chain, we also incorporated hydrodynamic interactions between beads i and j via the 3×3 matrix D_{ij} (which is itself a block of the $3N_b \times 3N_b$ diffusion tensor D). This matrix is given by

$$\mathbf{D}_{ij} = k_BT\left(\frac{\delta_{ij}}{\zeta}\mathbf{I} + (1-\delta_{ij})\mathbf{\Omega}_{ij}\right), \tag{6}$$

where δ_{ij} is the Kronecker delta function, $\zeta = 6\pi\eta a$ is the friction coefficient of a bead of hydrodynamic radius a in a solvent of viscosity η, $\mathbf{I}$ is an identity tensor, and $\mathbf{\Omega}_{ij}$ is the HI tensor, which relates the velocity perturbation at bead i to a point force at bead j. We computed $\mathbf{\Omega}_{ij}$ following the approach by Jendrejack *et al.*,[10] which is also described in detail elsewhere.[35,36] In the original implementation by Jendrejack *et al.*,[10] the wall correction to HI was calculated using the finite-element method. We have used the finite-difference method instead.[13]

Jendrejack *et al.*[32] parameterized this model based on free solution data for the relaxation time, diffusivity, and equilibrium stretch for λ-phage DNA at room temperature in a 43.3 cP solvent. We use their fitted parameters $b_k = 0.106\,\mu$m, $a = 0.077\,\mu$m, and $\phi = b_k^3 = 0.0012\,\mu\mathrm{m}^3$ in our simulations.

B. Brownian dynamics simulation algorithm

The motion of a bead-spring chain is governed by the Itô-Euler stochastic differential equation[37]

$$d\mathbf{R} = \left[\frac{1}{k_BT}\mathbf{D}\cdot\mathbf{F} + \frac{\partial}{\partial\mathbf{R}}\cdot\mathbf{D}\right]\Delta t + \sqrt{2}\mathbf{Br}\cdot d\mathbf{W}, \tag{7}$$

where $\mathbf{R}$ is a vector containing the $3N_b$ coordinates of the beads that constitute the DNA chain. The $3N_b \times 1$ vector $\mathbf{F}$ is the sum of the forces embodied by Eqs. (3)–(5), with $\mathbf{F}_i$ denoting the force vector acting on bead i. The components of the Gaussian noise $d\mathbf{W}$ are obtained from a real-valued Gaussian distribution with zero mean and variance Δt (the time step). The quantity $\mathbf{Br}$ is a tensor whose presence leads to multiplicative noise.[38,39] Its evaluation requires the decomposition of the diffusion tensor using the fluctuation-dissipation theorem, $\mathbf{D} = \mathbf{Br}\cdot\mathbf{Br}^T$. Note that while the divergence term $\partial/\partial\mathbf{R}\cdot\mathbf{D}$ vanishes in free solution, it is non-zero in confinement.

We made all equations dimensionless using the length scale a, time scale $\zeta a^2/k_BT$, and energy scale k_BT. Previous Brownian dynamics simulations of DNA in square channels with widths ranging from a few microns to hundreds of nanometers[10,35] were carried out by calculating both the stochastic term (noise term) and the divergence term in a matrix-free manner.[40–42] However, we do not use matrix-free approaches here because there is no computational

advantage —we need to store the diffusion tensor to calculate the Kirkwood diffusivity, so we can also use it in the integration of Eq. (7). We calculate the stochastic term using the Krylov subspace method.[43] The divergence term is not calculated explicitly but handled using mid-point algorithm.[40,42] The CPU time of the resulting algorithm scales as $\mathcal{O}(N_b^{2.25})$. This algorithm is useful for relatively small systems and has a favorable prefactor, but the N_b scaling eventually becomes unfavorable for a large number of beads.

We used a (dimensionless) time step of $\Delta t = 0.005$, which is an order of magnitude smaller than that used in Ref. 10. This small time step is important to capture the short-time behavior of the correlation functions described in Sec. III, which are the dominant contribution to the integrals required to compute the long-time diffusivity. We simulated molecular weights from $N_b = 6$ to 41 ($L = 10.5\,\mu$m to $84\,\mu$m) in dimensionless channel sizes $H = 6$ to 50 (462 nm to 3.85 μm). All simulations were run for up to 3.25×10^5 reduced time units. This corresponds to approximately 680 s, which is well within the range of typical experimental protocols. In order to obtain better statistics, we averaged over 24 runs (trajectories) for each set of parameters.

III. AXIAL DIFFUSIVITY OF A DNA CHAIN

We are interested in the axial diffusivity of DNA chain in square channels. This is a one-dimensional process due to the impenetrable channel walls. The Einstein relation in Eq. (2) for the one-dimensional diffusion along the axial (x) direction of the channel is

$$\langle \Delta R_{\mathrm{cm,x}}^2 \rangle = 2Dt, \tag{8}$$

where $\Delta R_{\mathrm{cm,x}}$ is the displacement of the x-component of the center-of-mass at some time t. Since $\langle \Delta R_{\mathrm{cm,x}}^2 \rangle$ for a polymer does not necessarily grow linearly with time for all t, then $D = D(t)$ with

$$D^{(K)} \equiv \lim_{t \to 0} \frac{\langle \Delta R_{\mathrm{cm,x}}^2 \rangle}{2t} \tag{9}$$

and

$$D_L \equiv \lim_{t \to \infty} \frac{\langle \Delta R_{\mathrm{cm,x}}^2 \rangle}{2t}. \tag{10}$$

An expression for D_L for a polymer chain in a channel can be derived by modifying the approach developed by Liu and Dünweg[24] in the context of free solution. The key modification in their approach for channel confinement is to account for the additional gradient term in Eq. (7). The detailed derivation, which follows in a straightforward manner from the work of Liu and Dünweg,[24] is presented in the supplementary material.[44] The key result is the final expression for D_L,[45]

$$D_L = D^{(K)} + \sum_{\Theta} \sum_{\Phi} D_{\Theta\phi} \qquad (\Theta = A, B, C \text{ and } \Phi = A, C), \tag{11}$$

where

$$D_{\Theta\Phi} = \frac{z}{N_b^2} \int_0^\infty \langle \Theta_x(0) \Phi_x(t) \rangle dt, \tag{12}$$

with

$$z = \begin{cases} (\Delta t)^{-1} & \text{if } \Theta = B \text{ and } \Phi = B \\ (\Delta t)^{-1/2} & \text{if } (\Theta \neq B \text{ and } \Phi = B) \text{ or } (\Theta = B \text{ and } \Phi \neq B) \\ 1 & \text{if } \Theta \neq B \text{ and } \Phi \neq B \end{cases}. \tag{13}$$

In Eq. (12), A, B, C correspond to the x-components of the different vector terms in Eq. (7) as written below

$$A_x = \left[\sum_{ij} \mathbf{D}_{ij} \cdot \mathbf{F}_j\right]_x,$$
$$B_x = \left[\frac{1}{\sqrt{\Delta t}} \sum_i \boldsymbol{\rho}_i\right]_x, \tag{14}$$
$$C_x = \left[\sum_{ij} \frac{\partial}{\partial \mathbf{R}_j} \cdot \mathbf{D}_{ij}\right]_x,$$

with $\boldsymbol{\rho}_i = \sum_j \sqrt{2}\, \mathbf{Br}_{ij} \cdot \Delta \mathbf{W}_j$. For notational simplicity, we suppress the subscript x for terms such as D_{AA}.[45]

IV. RESULTS

We have obtained data for four different molecular weights ($N_b = 6$, 11, 21, 41) in 10 different channel sizes ($H = 6$, 7, 8, 9, 10, 12, 15, 20, 30, 50). These 40 data sets exhibit many similar trends, so the basic phenomena are readily explained by considering a single parameter set. We thus begin here with the data obtained for $N_b = 11$ and $H = 10$, with the understanding that the qualitative conclusions drawn from this data set apply to the other channel sizes and molecular weights. We then consider the aggregate data to address our original goal concerning the accuracy of the Kirkwood approximation for describing experimental data for DNA in channel confinement.

A. Case study ($N_b = 11$ and $H = 10$)

In order to compute D_L through Eq. (11), we need to compute the correlations between the quantities A_x, B_x and C_x appearing in Eq. (14). These quantities are readily obtained during the integration of Eq. (7). Before proceeding to examining the correlations themselves, it is illuminating to examine first the order of magnitude of the different terms and their fluctuations. As an illustrative example, Fig. 1 presents these data for 100 reduced time units from a single trajectory. The values of A_x, B_x and C_x fluctuate around zero, but the magnitude of their fluctuations varies widely. The fluctuations in B_x are the largest, followed by A_x and then C_x. The results in Fig. 1 suggest that all of the correlation functions involving C_x will be very small. Figure 2 shows that this is indeed the case, with the four correlation functions involving the term C_x being very small in comparison to those excluding the term C_x. This result can be understood in terms of the variance σ^2 of the different quantities. For example, the particular

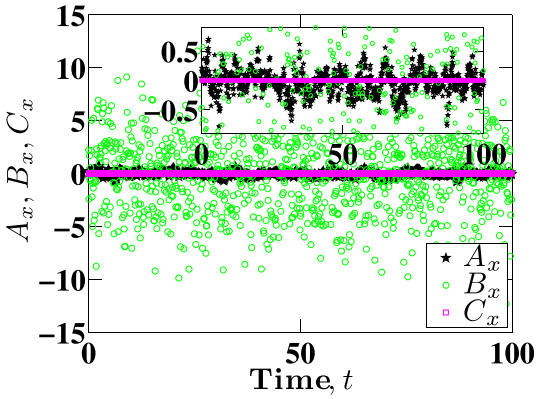

FIG. 1. Relative magnitudes of A_x, B_x, and C_x as functions of time for $N_b = 11$ and $H = 10$. The figure in the inset is the zoomed-in view to show that the magnitude of C_x is much smaller than that of A_x.

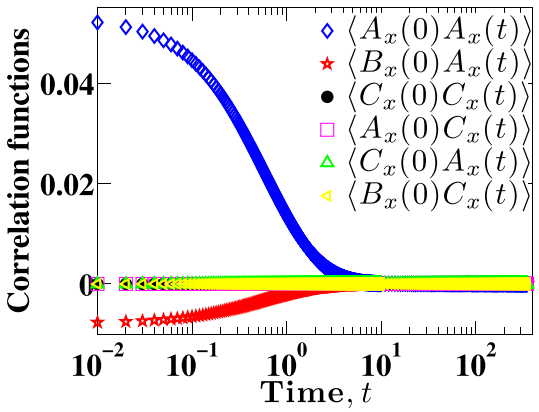

FIG. 2. Mean correlation functions of an 11 bead chain as functions of time in a channel of size $H = 10$.

subset of data in Fig. 1 leads to $\sigma^2_{B_x} = 13.9$, $\sigma^2_{A_x} = 0.07$, and $\sigma^2_{C_x} = 5 \times 10^{-6}$. As a result, we will assume that all contributions involving C_x make a negligible contribution to D_L. In other words, we conclude that the drift created due to the monomer concentration gradient, although non-zero in confinement, plays little role in the long-time diffusivity. Note that this does not mean that we can neglect the $\partial/\partial \mathbf{R} \cdot \mathbf{D}$ term when integrating Eq. (7), since that would lead to unphysical drift of the center-of-mass during the dynamic simulation. To compute D_L, we need to integrate correlation functions such as those appearing in Fig. 2. We use 64-point Gaussian quadrature to calculate these integrals numerically. Note that the total data set for a given correlation function typically contain around 10^5 points, so this quadrature is still very small and allows us to capture the details of the tail of the correlation function. To test for the convergence of integrals in Eq. (12), we calculate the cumulative integrals,

$$\tilde{D}_{\Theta\Phi}(t) = \frac{z}{N_b^2}\int_0^t \langle \Theta_x(0)\Phi_x(\tau)\rangle d\tau, \tag{15}$$

for different upper bounds, t. The resulting cumulative integrals for all correction terms are shown in Fig. 3. In an ideal case,

$$\lim_{t\to\infty} \tilde{D}_{\Theta\Phi}(t) = D_{\Theta\Phi}. \tag{16}$$

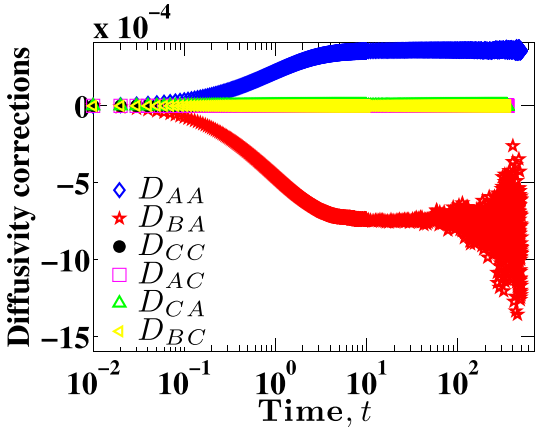

FIG. 3. Six different diffusivity correction terms D_{AA}, D_{BA}, D_{CC}, D_{AC}, D_{CA}, D_{BC} for $N_b = 11$ and for $H = 10$ as functions of time.

However, the data in the tails of the correlation become noisier as time-lag increases because the number of sample points decreases. Nevertheless, $\tilde{D}_{\Theta\Phi}(t)$ reaches a stable value for some large t before the sampling error becomes substantial. As expected from the correlation data shown in Fig. 2, the correction terms D_{CC}, D_{AC}, D_{CA}, and D_{BC} are negligible compared to D_{AA} and D_{BA}. We thus conclude that Eq. (11) can be written as

$$D_L \approx D^{(K)} + D_{AA} + D_{BA}. \tag{17}$$

The only quantity that remains to be calculated is the Kirkwood diffusivity. The Kirkwood diffusivity is calculated using one-dimensional form of Eq. (1) or Eq. (S-7) of the supplementary material,[44] where D_{ij} is generated for the configuration in each time step using Eq. (6). For this case study, we found that $D^{(K)} = 0.05883 \pm 0.00002$. Figure 4 shows how the diffusivity $D(t) = D^{(K)} + D_{AA}(t) + D_{BA}(t)$ depends on time. It is clear that $D^{(K)} > D(t \to \infty) = D_L$, and the overall shape of the curve is reminiscent of results obtained in free solution.[24]

B. Diffusivity corrections: Error analysis

Due to the finite number of data points available for long times, there is always sampling error in the tails of the correlation data in Fig. 2. As a result, we cannot estimate $D_{\Theta\Phi}$ by simply using $\tilde{D}_{\Theta\Phi}(t_{\max})$, where $t_{\max}$ is the maximum value of the simulation time. In order to arrive at a reasonable estimate for $D_{\Theta\Phi}$ and its uncertainty, we average over a subset of the data

$$D_{\Theta\Phi} = \frac{1}{N_{\text{sub}}} \sum_{j=0}^{N_{\text{sub}}} \tilde{D}_{\Theta\Phi}\left(t^l + j\Delta t_c\right), \tag{18}$$

where $N_{\text{sub}} = (t^u - t^l)/\Delta t_c$ is the number of data points between some lower bound t^l and upper bound t^u and $\Delta t_c = 0.01$ is the interval used for computing correlation functions. The corresponding error in $D_{\Theta\Phi}$ is estimated by the method of Chodera *et al.*[46] from the time series of data in this interval. For a given data set, the lower bound t^l is decided visually by eye such that t^l is always greater than the time at which the values of diffusivity corrections level off. We use $t^u = 500$ as the upper bound for any data set for which there are at least 30 000 data points between t^l and t^u. Note that the time series of these 30 000 data points is not completely uncorrelated. Therefore, the effective number of uncorrelated data points[46] turns out to be of the order of 100 in our study, which is the source of the error bar on $D_{\Theta\Phi}$. The value of $t^u = 500$ excludes the poor sampling for long times (e.g., consider D_{BA} in Fig. 3) because the total simulation length is typically 3×10^5; a typical correlation function at $t = 500$ includes 10^8 data points. In some cases, when the convergence of diffusivity corrections was not achieved until $t^l = 500$, the upper bound was chosen up to be $t^u = 3000$. For instance, for the largest

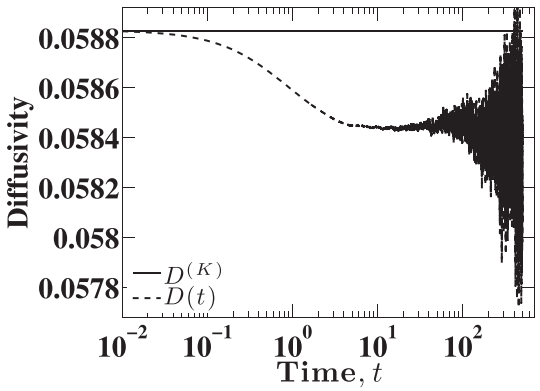

FIG. 4. Illustration of the drop in the diffusivity from its Kirkwood value to D_L for $N_b = 11$ and $H = 10$.

channel $H = 50$, we used $t^{\mu} = 2500$ for $N_b = 21$ and $t^{\mu} = 3000$ for $N_b = 41$. For the longest chains and largest channel sizes, the region between t^{l} and t^{μ} becomes relatively small, leading to the largest errors in $D_{\Theta\Phi}$. The correlation and diffusivity correction data for these cases, along with many other cases covering the range of parameters used in this work, are shown in the supplemental material.[44]

C. Convergence of D_L with respect to simulation time

As we noted in the introduction, reaching the asymptotic limit for D_L implied by Eq. (2) may require extremely long simulations. To quantify this statement, Fig. 5 shows the duration of our BD simulations, T, relative to the center-of-mass diffusion time $\tau \approx \langle R_g^2 \rangle / D^{(K)}$ for a particular molecular weight and channel size. In the latter, $\langle R_g^2 \rangle$ is the mean-square radius of gyration, which we use instead of the mean span of the confined DNA molecule to be consistent with prior work in free solution.[24] Simulations in free solution[24] suggest that a value $T/\tau \approx 10^2$ is required to obtain the asymptotic limit for D_L. For the smaller chains and larger channels, our simulations appear to reach this heuristic for the simulation time. For longer chains and smaller channels, it is infeasible to run simulations for $T/\tau \approx 10^2$. For example, consider the worst case scenario $N_b = 41$ and $H = 6$. For this case, we obtain the diffusion time $\tau \approx 2.41 \times 10^5$. We used one week of CPU time to run the BD simulation for $T = 3.25 \times 10^5$, roughly the same as the diffusion time. To achieve $T/\tau \approx 100$ would require two years to obtain a single data point. While such a calculation may be of theoretical interest, it is not particularly interesting in practice—as we noted in the introduction, a simulation time of $T = 3.25 \times 10^5$ corresponds to around 10 min and thus exceeds a typical image acquisition time in experiments. Moreover, it is not entirely clear that the heuristic limit $T/\tau \approx 10^2$ to obtain D_L in free solution[24] will hold in confinement. First, the configurational phase space in confinement is considerably smaller than in free solution. Second, and more importantly, the solid walls provide hydrodynamic screening and thus remove correlations between different parts of chain. Both of these factors suggest that the asymptotic limit may be reached more quickly in confinement than in free solution even after taking into account the slowing of the dynamics by polymer-wall friction.

Since there is no *a priori* argument to suggest a good heuristic for T/τ in confinement, we investigated the convergence of the value of D_L as a function of simulation time T for the smallest channel $H = 6$, which we presume will be the slowest to converge. Figures 6(a) and 6(c) plot the values of $D(t)$ obtained using four different values of T for the shortest chain ($N_b = 6$) and the longest chain ($N_b = 41$), respectively. The data in these figures, which are analogous to Fig. 4, suggest that the diffusivity is independent of the upper bound T for sufficiently large T. To be more quantitative, Figs. 6(b) and 6(d) plot the values of D_L obtained for both molecular weights, respectively, as a function of simulation time T. It is clear that the difference between mean values of D_L for the two longest times is not statistically significant, having p-values of 0.834 for $N_b = 6$ and 0.935 for $N_b = 41$. Table I shows the p-values for all pairs of

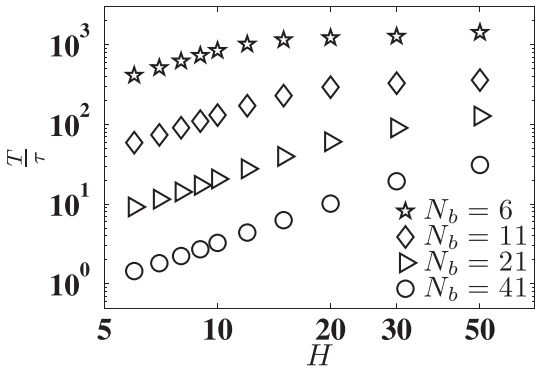

FIG. 5. Length of BD simulation T relative to the center-of-mass diffusion time of the chain τ.

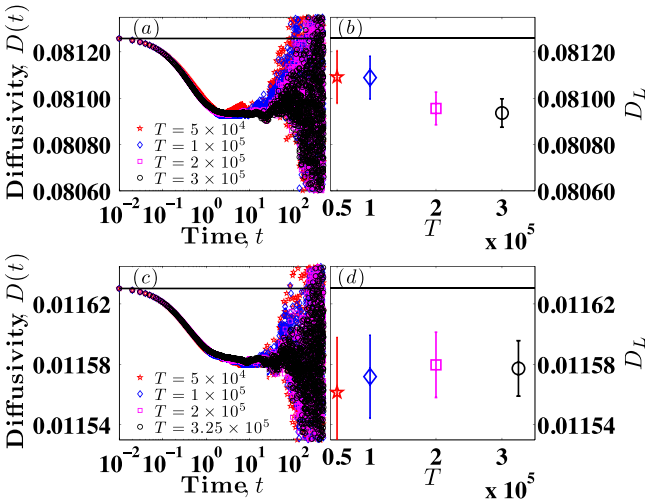

FIG. 6. Convergence of D_L for $N_b = 6$ and 41. (a) Time series $D(t)$ based on different simulation lengths for $N_b = 6$, $H = 6$. (b) Mean value of D_L as a function of simulation length T for $N_b = 6$, $H = 6$. (c) Time series $D(t)$ based on different simulation lengths for $N_b = 41$, $H = 6$. (d) Mean value of D_L as a function of simulation length T for $N_b = 41$, $H = 6$. Solid black lines in all four subplots indicate corresponding values of $D^{(K)}$ for that molecular weight.

TABLE I. p-values for various pairs of simulations lengths for both $N_b = 6$ (upper triangular matrix highlighted by italic cells) and $N_b = 41$ (lower triangular matrix highlighted by boldface cells). $T^\star$ denotes that it is different for both chains; for $N_b = 6$, $x = 3$ and for $N_b = 41$, $x = 3.25$.

	$T = 5 \times 10^4$	$T = 1 \times 10^5$	$T = 2 \times 10^5$	$T^\star = x \times 10^5$
$T = 5 \times 10^4$	1	*0.9862*	*0.3163*	*0.2356*
$T = 1 \times 10^5$	**0.8204**	1	*0.2581*	*0.1748*
$T = 2 \times 10^5$	**0.6691**	**0.8243**	1	*0.8340*
$T^\star = x \times 10^5$	**0.6988**	**0.8693**	**0.9345**	1

simulation lengths. Note that while calculating the p-values, the degree of freedom was 32 (effective number of uncorrelated data points) in these cases.

D. Effect of molecular weight and channel size

We repeated the analysis described in Sec. IV A for $N_b = 11$ and $H = 10$ for the other 39 combinations of molecular weight and channel size. We find the same trend for all sets of parameters, namely, that there are only two dominating diffusivity correction terms D_{AA} and D_{BA}, and the others are negligible. Plots of correlation functions and diffusivity corrections for some of the parameters sets are shown in the supplementary material[44] to show the generality of this trend. We also include a table of Kirkwood diffusivity values for each combination of N_b and H in the supplementary material.[44] These data allow one to produce the equivalent of Fig. 4 for different channel sizes and molecular weights through Eq. (17).

A careful inspection of Fig. 3 suggests that $D_{BA} \approx -2D_{AA}$, which is also the case from Brownian dynamics simulations in free solution.[24] Figure 7 shows that this is indeed the case for all channel sizes and molecular weights, except for channel size $H = 50$. The data for $H = 50$, which correspond to weak confinement for all of the chain sizes considered here, have relatively larger error bars for all chains. We cannot confirm $D_{BA} \approx -2D_{AA}$ for this channel. However, for all other parameters, we can infer that

$$2D_{AA} + D_{BA} \approx 0. \tag{19}$$

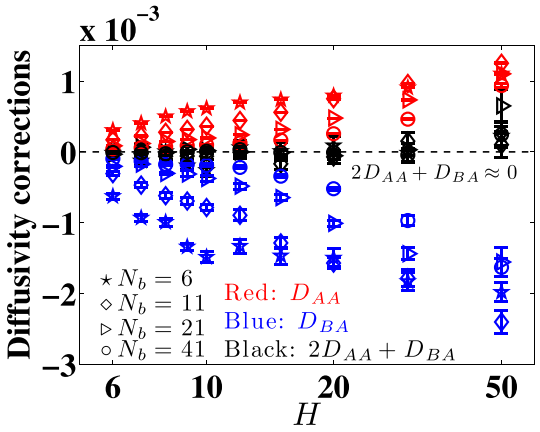

FIG. 7. Converged values of D_{AA} and D_{BA} for a range of channels sizes and chain sizes. The black data points, except for $H = 50$, show that $D_{BA} \approx -2D_{AA}$.

We emphasize that this is not proof of the equality $D_{BA} = -2D_{AA}$, but rather data suggesting that such an equality likely holds in confinement.

The corrections to the diffusion coefficient in channel confinement are essentially the same as those found by Liu and Dünweg[24] in Brownian dynamics simulations of a polymer chain in free solution. Note, however, a subtle difference in methodology; the Kirkwood diffusivity in free solution is calculated using Eq. (1), not the one-dimensional form of Eq. (1) or Eq. (S-7) of the supplementary material,[44] and the correction terms are calculated using the 3D version of Eqs. (11)–(13), where D_{AA} and D_{BA} are given by

$$\begin{aligned} D_{AA} &= \frac{1}{3N_b^2}\int_0^\infty \langle \mathbf{A}(0)\cdot\mathbf{A}(t)\rangle dt, \\ D_{BA} &= \frac{1}{3N_b^2\sqrt{\Delta t}}\int_0^\infty \langle \mathbf{B}(0)\cdot\mathbf{A}(t)\rangle dt, \end{aligned} \tag{20}$$

with $\mathbf{A}$ and $\mathbf{B}$ being the vectors defined in Eq. (14).

As a result of Eq. (19), Eq. (17) can further be simplified to

$$D_L \approx D^{(K)} - D_{AA}, \tag{21}$$

which is exactly the same relationship derived by Fixman[18] for free solution for long-time diffusivity in the theoretical limit. Moreover, since D_{AA} is positive in confinement, we further find that

$$D_L < D^{(K)}, \tag{22}$$

again in agreement with Fixman's results for free solution for long-time diffusivity in the theoretical limit.[18]

We are now at the stage where we can achieve our original goal of assessing the accuracy of the Kirkwood approximation in channel-confinement. Following the standard convention in the literature,[18,20] we define the quantity

$$\Delta D \equiv 100 \times \frac{D^{(K)} - D_L}{D^{(K)}} \tag{23}$$

as the percent error due to Kirkwood approximation. Using Eq. (17), we can also express this deviation in terms of the significant correction terms,

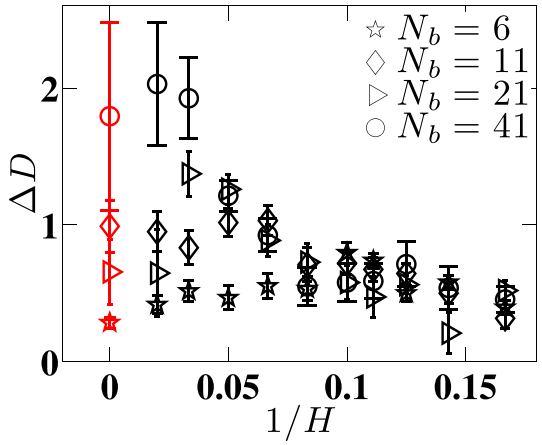

FIG. 8. ΔD for a range of channel sizes and chain sizes. Note that the red data for $1/H = 0$ correspond to free solution data.

$$\Delta D \approx 100 \times \frac{-(D_{AA} + D_{BA})}{D^{(K)}}. \tag{24}$$

Figure 8 presents the error ΔD in the Kirkwood approximation for all of the simulations included in our analysis. The error is always less than about 2%.

V. DISCUSSION

Many of the results for channel-confined DNA mimic those in free solution, most notably Eqs. (19) and (22) and the negligible contribution to D_L from C_x (which is identically zero in free solution). We also observe an increase in the error for the Kirkwood approximation in free solution as the molecular weight increases. Similar to Liu and Dünweg,[24] we cannot reach sufficiently high molecular weights to determine whether the error in the Kirkwood diffusivity for this DNA model continues to increase as N_b increases further or whether it plateaus at some point. Overall, the relatively low error of the Kirkwood approximation for DNA in free solution is encouraging because we have used the Kirkwood approximation elsewhere[47] to make broader conclusions about the utility of DNA as a model polymer. An important feature of the key result in Fig. 8 is the collapse of the data for the smaller channel sizes. To understand the collapse, we note that the data in this figure correspond to three cases: (i) free solution ($H \to \infty$), (ii) weak confinement ($H \gtrsim R_g$), and (iii) the de Gennes regime. To distinguish more quantitative between cases (ii) and (iii), Fig. 9 plots the mean span,

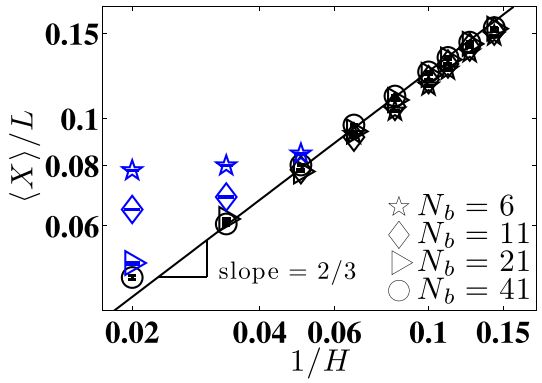

FIG. 9. The ratio of mean extension $\langle X \rangle$ to contour length L of the chain as a function of channel size H for four different chain sizes. The solid line is Eq. (26) for the de Gennes scaling for the extension of a chain in a nanochannel. The blue symbols are weak confinement, and the black symbols are those chains that are long enough to be in the de Gennes regime.

$$\langle X \rangle = \langle \max(R_{i,x}) - \min(R_{i,x}) \rangle, \tag{25}$$

of the chain as a function of the inverse channel size. In the latter, $R_{i,x}$ is the x component of the position vector $\mathbf{R}_i$ of bead i. Those data lying on the de Gennes scaling law,[28]

$$\langle X \rangle \sim H^{-2/3}, \tag{26}$$

correspond to confinement in the de Gennes regime. The remaining data represent weak confinement; the DNA experiences hydrodynamic (and excluded volume) interactions with the walls, but it has not yet achieved the blob-like configuration that characterizes confinement in the de Gennes regime because the molecular weight is too low.

Once the chains are long enough to be in the de Gennes regime, we can understand the error in the Kirkwood approximation (as well as the collapse of the data in Fig. 8) through the blob hydrodynamics ideas of Brochard and de Gennes.[9] In blob theory, each subchain of size

$$H \sim L_{\text{blob}}^{3/5} l_p^{1/5} w^{1/5} \tag{27}$$

is envisioned as a swollen excluded volume subchain of length L_{blob} that is hydrodynamically decoupled from the other blobs through the screening provided by the walls. As a result, the properties of the chain are readily computed by determining the properties of a single blob and then summing up the contribution due to the

$$N_{\text{blob}} = L/L_{\text{blob}}, \tag{28}$$

blobs in the chain. For example, blob theory leads to the scaling in Eq. (26) for the chain extension by noting that span of the confined chain is $\langle X \rangle \sim N_{\text{blob}} H$.

In blob theory, the diffusivity of the chain with hydrodynamically independent blobs is given by[9]

$$D \approx \frac{k_B T}{N_{\text{blob}} \zeta_{\text{blob}}} = \frac{D_{\text{blob}}}{N_{\text{blob}}}, \tag{29}$$

where ζ_{blob} is the friction coefficient of a blob and D_{blob} is the corresponding diffusivity of a blob. Naturally, we would expect that the friction coefficient of a blob in the short-time limit will differ from that in the longer time limit. If we denote these diffusivities as $D_{\text{blob}}^{(K)}$ and $D_{\text{blob},L}$, respectively, and insert Eq. (29) in the definition of ΔD in Eq. (23), we arrive at

$$\Delta D \equiv 100 \times \frac{D_{\text{blob}}^{(K)} - D_{\text{blob},L}}{D_{\text{blob}}^{(K)}}. \tag{30}$$

The latter result is independent of the length of the chain, which explains the collapse of the data in Fig. 8. When we combine Eq. (30) with the dependence of ΔD on N_b in free solution in Fig. 8 and prior work,[24] we can also rationalize the drop in ΔD with increasing confinement in the de Gennes regime. As we reduce H, Eq. (27) indicates that the length of the subchain comprising that blob is reduced as well. Since the error in the Kirkwood approximation decreases with decrease in molecular weight in free solution for any of the blob sizes here (since ΔD decreases for all of the chain sizes N_b that we studied and $N_b > N_{\text{blob}}$ by definition), blob theory implies that the error in the Kirkwood approximation also decreases with the decrease in blob size.

The mean values of ΔD in Fig. 8 for chains that are not in the de Gennes regime suggest that there might be a maximum in ΔD as channel size increases and approaches infinity (free solution case). However, due to relatively large error bars on ΔD in the weak confinement regime, we cannot confirm such a peak using our simulation results. However, if there really exists such a maximum or if ΔD decreases as channel size increases beyond that required to

have a single blob, it might be due to the reduction in the frequency of the monomer-wall hydrodynamic interactions, which is one of the source of the error due to Kirkwood approximation (other than monomer-monomer HI).

VI. CONCLUDING REMARKS

Using a Brownian dynamics simulation algorithm with excluded volume and hydrodynamic interactions (HI),[10] we have shown that the error in the Kirkwood diffusivity decreases with the increase in confinement in the de Gennes regime due to the screening of hydrodynamic interactions between blobs. Overall, the error in the Kirkwood approximation is rather small for this DNA model, peaking at around 2% in weak confinement for a relatively large chain. Moreover, the error becomes independent of chain length once the polymer is large enough to be confined by the channel. These results lend support to conclusions about the diffusivity of DNA drawn from the Kirkwood approximation in both free solution[47] and confinement.[13,17]

A key open question concerns how the error in the Kirkwood approximation changes as confinement increases further. In particular, we are motivated in our work by genome mapping in nanochannels,[4] which takes place in channel sizes close to the 50 nm persistence length of DNA, an order of magnitude smaller than the smallest channel size considered here. As the channel size approaches the persistence length, the de Gennes blob theory breaks down and it is not obvious whether hydrodynamic conclusions drawn from a blob model apply to even stronger confinement. However, we have reason to be optimistic that the error in stronger confinement will be no worse (and probably better) than in the de Gennes regime. The error in the Kirkwood approximation arises from correlations in the monomer-monomer and monomer-wall hydrodynamic interactions. As we increase the confinement further, there is a decrease in both (i) the number of monomers inside the hydrodynamic screening volume H^3 and (ii) the range of possible distances between a monomer and the wall. As a result, we would expect that the correlations in monomer-monomer HI and monomer-wall HI will both decrease; in the former case, due to a reduction in the number of monomer-monomer pairs, in the latter case, due to a reduction in the magnitude of the fluctuations in monomer-wall HI.

Testing this hypothesis about the accuracy of the Kirkwood approximation in strong confinement is not a simple task. The coarse-grained DNA model that we used here, which has a discretization of almost 40 persistence lengths per spring, is only valid in the de Gennes regime. This model breaks down in stronger confinement because it is unable to resolve the chain configurations at the length scale of a persistence length. There are other options to model polymers in such strong confinement, such as the touching-bead model[48] that we have used previously in our studies of the Kirkwood diffusivity of DNA in confinement.[13] Unfortunately, dynamic simulations of the touching-bead model would likely require simulating thousands of beads to reach the long-chain limit. Moreover, it seems likely that the relaxation time scales quadratically with molecular weight, making such simulations extremely expensive.

Ultimately, we suspect that detailed simulation of the dynamics of DNA confined in very small channels is not necessary to model the practical circumstances of interest such as genome mapping. Our results indicate that the Kirkwood approximation is quite accurate for DNA in the accessible regimes of confinement for a simulation. Moreover, it seems unlikely that there would be a sharp increase in the error of the Kirkwood approximation when confinement is increased further. Given the inherent uncertainty in any experiment relative to the error in the Kirkwood approximation, we conclude that the Kirkwood approximation is a useful tradeoff between simulation cost and time for modeling the diffusivity of confined DNA.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health (R01-HG006851). Computational resources were provided in part by the University of Minnesota Supercomputing Institute.

[1]Z. T. Berndsen, N. Keller, S. Grimes, P. J. Jardine, and D. E. Smith, Proc. Natl. Acad. Sci. USA **111**, 8345 (2014).
[2]K. D. Dorfman, Rev. Mod. Phys. **82**, 2903 (2010).

[3]J. Tegenfeldt, C. Prinz, H. Cao, S. Chou, W. Reisner, R. Riehn, Y. M. Wang, E. C. Cox, J. C. Sturm, P. Silberzan, and R. H. Austin, Proc. Natl. Acad. Sci. USA **101**, 10979 (2004).
[4]E. T. Lam, A. Hastie, C. Lin, D. Ehrlich, S. K. Das, M. D. Austin, P. Deshpande, H. Cao, N. Nagarajan, M. Xiao, and P.-Y. Kwok, Nat. Biotechnol. **30**, 771 (2012).
[5]R. B. Bird, C. F. Curtiss, R. C. Armstrong, and O. Hassager, *Dynamics of Polymeric Liquids*, Kinetic Theory Vol. 2 (John Wiley & Sons, New York, 1986).
[6]A. Karpusenko, J. H. Carpenter, C. Zhou, S. F. Lim, and R. Riehn, J. Appl. Phys. **111**, 024701 (2012).
[7]D. R. Tree, Y. Wang, and K. D. Dorfman, Biomicrofluidics **7**, 054118 (2013).
[8]M. D. Graham, Annu. Rev. Fluid Mech. **43**, 273 (2011).
[9]F. Brochard and P. G. de Gennes, J. Chem. Phys. **67**, 52 (1977).
[10]R. M. Jendrejack, D. C. Schwartz, M. D. Graham, and J. J. de Pablo, J. Chem. Phys. **119**, 1165 (2003).
[11]Y.-L. Chen, Biomicrofluidics **7**, 054119 (2013).
[12]Y.-L. Chen, Y. Lin, J.-F. Chang, and P.-K. Lin, Macromolecules **47**, 1199 (2014).
[13]D. R. Tree, Y. Wang, and K. D. Dorfman, Phys. Rev. Lett. **108**, 228105 (2012).
[14]J. G. Kirkwood and J. Riseman, J. Chem. Phys. **16**, 565 (1948).
[15]J. G. Kirkwood, J. Polym. Sci. **12**, 1 (1954).
[16]D. R. Tree, Y. Wang, and K. D. Dorfman, Phys. Rev. Lett. **110**, 208103 (2013).
[17]L. Dai, D. R. Tree, J. R. C. van der Maarel, K. D. Dorfman, and P. S. Doyle, Phys. Rev. Lett. **110**, 168105 (2013).
[18]M. Fixman, Macromolecules **14**, 1710 (1981).
[19]M. Fixman, J. Chem. Phys. **78**, 1594 (1983).
[20]M. Fixman, J. Chem. Phys. **84**, 4080 (1986).
[21]B. H. Zimm, Macromolecules **13**, 592 (1980).
[22]J. Garcia de la Torre, A. Jimenez, and J. J. Freire, Macromolecules **15**, 148 (1982).
[23]S. Q. Wang, J. F. Douglas, and K. F. Freed, J. Chem. Phys. **85**, 3674 (1986).
[24]B. Liu and B. Dünweg, J. Chem. Phys. **118**, 8061 (2003).
[25]M. L. Mansfield, J. F. Douglas, S. Irfan, and E.-H. Kang, Macromolecules **40**, 2575 (2007).
[26]R. Schmidt, J. Cifre, and J. Garcia de la Torre, Eur. Phys. J. E **35**, 130 (2012).
[27]W. Reisner, K. J. Morton, R. Riehn, Y. M. Wang, Z. Yu, M. Rosen, J. C. Sturm, S. Y. Chou, E. Frey, and R. H. Austin, Phys. Rev. Lett. **94**, 196101 (2005).
[28]M. Daoud and P. G. de Gennes, J. Phys. France **38**, 85 (1977).
[29]T. Odijk, Phys. Rev. E **77**, 060901(R) (2008).
[30]Y. Wang, D. R. Tree, and K. D. Dorfman, Macromolecules **44**, 6594 (2011).
[31]Y.-L. Chen, M. D. Graham, J. J. de Pablo, G. C. Randall, M. Gupta, and P. S. Doyle, Phys. Rev. E **70**, 060901(R) (2004).
[32]R. M. Jendrejack, J. J. de Pablo, and M. D. Graham, J. Chem. Phys. **116**, 7752 (2002).
[33]J. F. Marko and E. D. Siggia, Macromolecules **27**, 981 (1994).
[34]J. F. Marko and E. D. Siggia, Macromolecules **28**, 8759 (1995).
[35]R. M. Jendrejack, D. C. Schwartz, J. J. de Pablo, and M. D. Graham, J. Chem. Phys. **120**, 2513 (2004).
[36]K. D. Dorfman, D. Gupta, A. Jain, A. Muralidhar, and D. R. Tree, Eur. Phys. J. Spec. Top. **223**, 3179 (2014).
[37]D. L. Ermak and J. A. McCammon, J. Chem. Phys. **69**, 1352 (1978).
[38]H. C. Ottinger, *Stochastic Processes in Polymeric Fluids* (Springer, Berlin, 1996).
[39]In Eq. (7), we adopt the notation Br for the multiplicative noise, rather than the standard notation B, to avoid possible confusion with the symbol B_x used in the derivation of the long-time diffusion coefficient in Sec. III.
[40]M. Fixman, J. Chem. Phys. **69**, 1527 (1978).
[41]M. Fixman, Macromolecules **19**, 1204 (1986).
[42]P. S. Grassia, E. J. Hinch, and L. C. Nitsche, J. Fluid Mech. **282**, 373 (1995).
[43]T. Ando, E. Chow, Y. Saad, and J. Skolnick, J. Chem. Phys. **137**, 064106 (2012).
[44]See supplementary material at http://dx.doi.org/10.1063/1.4917269 for: (i) the derivation of the axial long-time diffusivity, (ii) autocorrelation and diffusivity corrections for $N_b = 6$, 11, 21, and 41 in channel sizes $H = 6$, 8, 15, and 50, and (iii) Kirkwood diffusivities for all channel sizes and molecular weights.
[45]In the literature on diffusion of polymers in free solution,[18,24] the quantities D_{AA} and D_{BA} have been denoted by D_1 and D_2, respectively. Since there are so many combinations of subscripts in confinement, we adopted the notation of Eq. (11) to make the subsequent text easier to follow.
[46]J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill, J. Chem. Theory Comput. **3**, 26 (2007).
[47]D. R. Tree, A. Muralidhar, P. S. Doyle, and K. D. Dorfman, Macromolecules **46**, 8369 (2013).
[48]J. Wang and H. Gao, J. Chem. Phys. **123**, 084906 (2005).