Bubble merging in breathing DNA as a vicious walker problem in opposite potentials

Pedersen, Jonas Nyvold; Hansen, Mikael Sonne; Novotny, Tomas; Ambjornsson, Tobias; Metzler, R.

Published in:
Journal of Chemical Physics

Link to article, DOI:
10.1063/1.3117922

Publication date:
2009

Document Version
Publisher's PDF, also known as Version of record

Bubble merging in breathing DNA as a vicious walker problem in opposite potentials

Jonas Nyvold Pedersen,[1] Mikael Sonne Hansen,[2] Tomáš Novotný,[3] Tobias Ambjörnsson,[4,a)] and Ralf Metzler[5,b)]
[1]*Department of Mathematical Physics, Lund University, Box 118, 22100 Lund, Sweden*
[2]*Department of Mathematics, Technical University of Denmark, Bldg. 303S, Matematiktorvet, 2800 Kongens Lyngby, Denmark*
[3]*Department of Condensed Matter Physics, Faculty of Mathematics and Physics, Charles University, Ke Karlovu 5, 121 16 Prague, Czech Republic and Nano-Science Center, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark*
[4]*Department of Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*
[5]*Department of Physics, Technical University of Munich, James Franck Straße, D-85747 Garching, Germany*

(Received 1 December 2008; accepted 23 March 2009; published online 30 April 2009)

We investigate the coalescence of two DNA bubbles initially located at weak domains and separated by a more stable barrier region in a designed construct of double-stranded DNA. In a continuum Fokker–Planck approach, the characteristic time for bubble coalescence and the corresponding distribution are derived, as well as the distribution of coalescence positions along the barrier. Below the melting temperature, we find a Kramers-type barrier crossing behavior, while at high temperatures, the bubble corners perform drift diffusion toward coalescence. In the calculations, we map the bubble dynamics on the problem of two vicious walkers in opposite potentials. We also present a discrete master equation approach to the bubble coalescence problem. Numerical evaluation and stochastic simulation of the master equation show excellent agreement with the results from the continuum approach. Given that the coalesced state is thermodynamically stabilized against a state where only one or a few of the base pairs of the barrier region are re-established, it appears likely that this type of setup could be useful for the quantitative investigation of thermodynamic DNA stability data as well as the rate constants involved in the unzipping and zipping dynamics of DNA in single molecule fluorescence experiments. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3117922]

I. INTRODUCTION

Within a broad range of salt and temperature conditions, the Watson–Crick double helix[1] is the equilibrium structure of DNA. This thermodynamic stability is effected by hydrogen bonding between paired bases and by base stacking between nearest neighbor pairs of base pairs (bps).[1–8] By an increase in the temperature or by variation of the *p*H value (titration with acid or alkali) double-stranded DNA progressively denatures, yielding regions of single-stranded DNA, until the double strand is fully molten. This is the helix-coil transition.[9,10] The melting temperature T_m is defined as the temperature at which one-half of the DNA molecule has undergone denaturation.[5,9,11,12] Typically, the denaturation starts in regions rich in the weaker *adenine-thymine* (AT) bps and subsequently moves to zones of increasing *guanine-cytosine* (GC) content. The occurrence of zones of different stabilities within the genome was shown to be relevant when separating coding from noncoding regions.[13,14]

However, already at room temperature thermal fluctuations cause rare opening events of small intermittent denaturation zones in the double helix.[15] These *DNA bubbles* consist of flexible single-stranded DNA, and their size fluctuates by stepwise zipping and unzipping of the bps at the zipper forks, where the bubble connects to the intact double strand. Initiation of a bubble in a stretch of intact double strand requires the crossing of a free energy barrier $\Delta G_{\mathrm{bubble}}$ of some $8k_BT-12k_BT$ at physiological temperature, corresponding to a Boltzmann factor, often referred to as the cooperativity factor, $\sigma_0=\exp(-\Delta G_{\mathrm{bubble}}/k_BT)\sim 10^{-5}-10^{-3}$. Once formed below the melting temperature T_m, a bubble will eventually zip close. Above T_m, a bubble will preferentially stay open and, if unconstrained, grow in size until it merges with other denaturation bubbles, eventually leading to full denaturation of the double helix. Constraints against such full unzipping could, for instance, be the buildup of twist in smaller DNA rings or the chemical connection of the two strands by short bulge loops, compare Ref. 16.

Biologically, the physical conformations of DNA molecules are considered of increasing relevance for its function, see, for instance, the review in Ref. 17 and references therein. In particular, the existence of intermittent (though infrequent) bubble domains is important, as the opening up of the Watson–Crick bps by breaking of the hydrogen bonds between complementary bases disrupts the helical stack. The

[a)]Present address: Department of Theoretical Physics, Lund University, Sölvegatan 14A, 223 62 Lund, Sweden.
[b)]Electronic mail: metz@ph.tum.de.

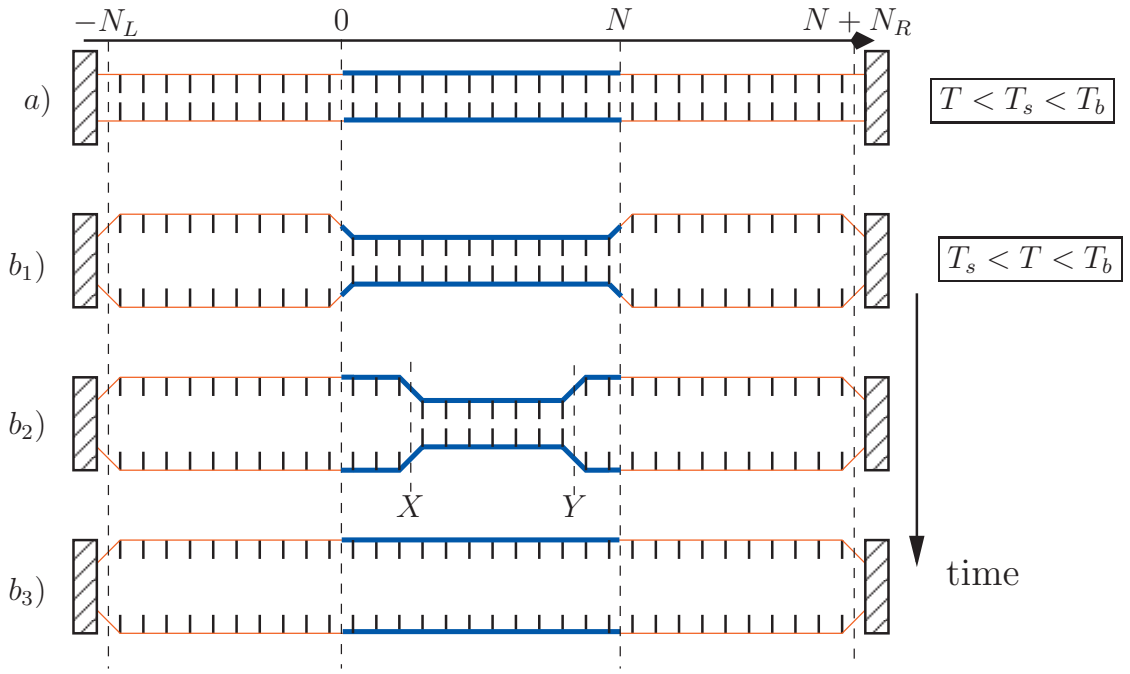

FIG. 1. (Color) Schematic of the bubble coalescence setup in a designed DNA construct. It is clamped at both ends and consists of two outer soft zones (thin red lines) of lengths N_L, N_R bps with melting temperature T_s and a stronger N-bps-long barrier zone (thick blue lines) with $T_b > T_s$. (a) All bps closed ($T < T_s < T_b$). (b) Soft zones open by raising the temperature above T_s. [$(b_1)-(b_3)$] Successive opening of the barrier driven mainly by fluctuations ($T < T_b$) or drift ($T > T_b$) until coalescence. The discrete coordinates $X, Y = -N_L, \ldots, N+N_R$ are defined as the positions of the *interfaces* between the closed and broken bps.

flipping out of the ordered stack of the unpaired bases allows the binding of specific chemicals or proteins, that otherwise would not be able to access the reactive sites of the bases.[5,8,9,15] In fact, there exists a competition of time scales between the opening/closing dynamics of DNA bubbles and the binding kinetics of selectively single-stranded DNA binding proteins.[18–21] That the chemical potential of the single-stranded binding proteins does not lead to full denaturation of the DNA is due to the slow binding of the proteins when compared to the bubble dynamics.[18–21] It is also believed that DNA breathing assists in the transcription initiation process.[22–25] The quantitative knowledge of the denaturation dynamics as well as energetics is imperative to a better understanding of genomic biochemical processes.

DNA breathing has been modeled extensively in terms of the Peyrard–Bishop model, that is based on a set of Langevin equations for the base-base distance in a bp; the effective attraction between the bases is represented by model potentials.[26–30] Alternatively, DNA breathing can be considered as a random walk process in the free energy landscape of the Poland–Scheraga model of DNA denaturation, as the number of broken bps turns out to be the slow variable of the process.[16] In continuum form, this approach to DNA bubble dynamics has been described in terms of a Fokker–Planck equation.[31–33] A discrete description, in which the coordinate of the random walker corresponds to a specific bp, was suggested in Refs. 21, 24, 34, and 35, and the corresponding stochastic simulation analysis of DNA breathing was introduced in Ref. 36. The influence of a random energy landscape on bubble localization and dynamics was studied in Refs. 32, 37, and 38, while a framework to include an arbitrary given sequence of bps was developed in Refs. 24, 25, and 39. Endowed with the sensitivity of their dynamics, DNA constructs were proposed as nanosensors.[40,41] We note that the formulation in terms of the gradient of the Poland–Scheraga free energy allows one to explicitly introduce all necessary independent stacking parameters based on the study in Ref. 8, see also the discussion in Ref. 25. Measuring the dynamics of DNA bubbles also provides information on the magnitude of the critical exponent c representing the entropy loss factor of a closed polymer loop, deciding on the order of the denaturation transition,[9,42–47] and thus also influencing the temporal survival probability of bubbles.[32,33]

The multistate nature of *DNA breathing* can be monitored in real time on the single DNA level by fluorescence correlation techniques.[16] It has been shown in a quantitative analysis that the experimentally accessible autocorrelation function is sensitive to the stacking parameters of DNA.[24,25] However, it has not been fully appreciated to what extent the fluorophore and quencher molecules, that are attached to the DNA construct in the experiments reported in Refs. 16, 48, and 49, influence the stability of DNA. Moreover, the zipping rates measured in the single molecule fluorescence setup differ from those determined in NMR experiments.[15,16] We here propose and study a complementary setup for the single molecule fluorescence investigation of DNA breathing, as shown in Fig. 1. In this setup, a short stretch of DNA, clamped at both ends, is designed such that two soft zones consisting of weaker AT bps are separated by a more stable barrier region rich in GC bps. For simplicity, we assume that both soft zones and barrier are homopolymers with bp-dissociation free energies ΔG_s and ΔG_b, respectively, and, in accordance with the experimental findings of Ref. 16, we neglect secondary structure formation in the barrier zone. At

temperatures higher than the melting temperature T_s of the soft zones but still lower than the melting temperature T_b of the barrier region such that two open bubbles are being promoted, thermal fluctuations will gradually dissociate the barrier until the two bubbles coalesce. Note that the melting temperature at 100 mM salt conditions differs by approximately 50 K between mixed $(AT/TA)_n$ and $(GC/CG)_n$ homopolymers, respectively.[7,8] This should provide a large enough temperature interval between hard and soft zones to perform this type of experiment. In the following we use realistic values for the simulations. Once coalesced, the free energy corresponding to one cooperativity factor $\sigma_0 \sim 10^{-5}-10^{-3}$ is released, stabilizing the coalesced bubble against reclosure of the barrier. This fact should allow for a meaningful measurement of the coalescence time in experiment, and therefore provide a new and sensitive method to measure DNA stability data and bp zipping rates. We also study the case when the system is prepared as above and then T suddenly increased such that $T>T_b>T_s$ so that the system is driven toward coalescence. In both cases the two boundaries between bubbles and barrier perform a (biased) random walk in opposite free energy potentials.

In fact, the study of the bubble coalescence is of interest in its own right, as we map the random walk of the two zipper forks separating double-stranded barrier bps from already denatured single-stranded bubble domains onto a new case of the vicious walker problem. Namely, we deal with two vicious walkers in *linear* but *opposite* potentials. The viciousness condition corresponds to the fact that when the two zipping forks meet, the bubbles coalesce and the dynamics is stopped. While the problem of a general number of (otherwise noninteracting) vicious walkers in free space was solved a long time ago[50] and has been only relatively recently generalized to the case of motion in a common potential,[51] even two walkers in different potentials cannot be addressed analogously by the straightforward antisymmetrization procedure. To solve our problem in the continuum limit, we use a trick of introducing individual symmetry transformations for each walker which transform the respective Fokker–Planck operators to the *same* Hermitian form. In the transformed frame the problem is solved by the standard procedure constructing the joint probability density from the antisymmetrized product of the single-walker probability densities. This solution enables us to effectively reduce the numerical efforts needed for the evaluation of the joint probability density, mean coalescence time, spatial probability density of coalescence position, etc., by one dimension. Moreover, some of the quantities of interest can be obtained this way fully analytically including all the characteristics of the high-barrier case which are very hard to reliably determine numerically.

The paper is organized as follows. After introducing the discrete model in Sec. II and its continuum limit in Sec. III, we solve the latter in Sec. IV by using the symmetry of the problem. The main quantities of interest, namely, the coalescence time and its distribution, as well as the distribution of the coalescence position, are obtained in Sec. V. In Sec. VI, we introduce a direct solution of the discrete problem via the complete master equation and a stochastic simulation scheme (Gillespie) and compare these results to those of the continuum approximation in Sec. VII. In Sec. VIII we address the connection between our models and results and real biologically relevant data. In Sec. IX we state our conclusions. Appendix A presents a detailed calculation of the auxiliary single-walker density. In Appendix B we explain the direct numerical solution to the full master equation of Sec. VI, while in Appendix C the Gillespie stochastic simulation scheme for the same master equation is briefly summarized.

II. DEFINING THE ZIPPING AND UNZIPPING RATES

In this section we define the transition rates for opening or closing a bp, which are determined by two effects: the energy landscape stemming from the thermodynamical partition factors and the thermal fluctuations.

As illustrated in Fig. 1, we consider the case when the two soft bubble zones of the DNA construct are preferentially open, while the central barrier region is initially completely closed. We assume that the two soft zones are homopolymers with identical melting temperature T_s, and that the barrier region is a homopolymer with melting temperature $T_b>T_s$. In the following, we neglect secondary structure formation in the bubbles, consistent with experimental observations in relatively short bubble domains.[16] The barrier region of initially closed bps between the zipper forks will also be referred to as the clamp.

Each of the two DNA bubbles is characterized by a partition factor of the form

$$\mathcal{Z}_L(X) = \frac{\sigma_0}{(X+N_L+1)^c} \prod_{\tilde{X}=-N_L}^{X} u(\tilde{X}), \tag{1}$$

$$\mathcal{Z}_R(Y) = \frac{\sigma_0}{(N+N_R-Y+1)^c} \prod_{\tilde{Y}=Y}^{N+N_R} u(\tilde{Y}), \tag{2}$$

where $X,Y \in [-N_L, N+N_R]$ denote the positions of the left and right zipper forks, respectively. In Eqs. (1) and (2) the quantity $u(X)$, which takes the values $u_{s(b)}$, is the Boltzmann factor for breaking a bp in the soft zone (barrier domain), $u_{s(b)}=\exp(\beta\Delta G_{s(b)})$, corresponding to the free energy $\Delta G_{s(b)}$ for breaking a bp; moreover, $\beta \equiv 1/[k_B T]$. We define $u(-N_L)=u(N+N_R)\equiv 1$.

Note once more the prefactor $[(\text{bubble length})+1]^{-c}$ displaying the inherent long-range character of the Poland–Scheraga free energy model. It measures the reduction of the degrees of freedom of a loop configuration, as characterized by the critical exponent c.[7,9,12,43,52] For the long-time behavior in larger, single bubbles the influence of c on the distribution of bubble lifetimes is considered in Refs. 32 and 33 in a continuum approach.

Finally, σ_0 is the cooperativity factor corresponding to the free energy barrier for breaking the first bp in a stretch of intact double strand. Loosely speaking, it corresponds to the disruption of two stacking interactions in the DNA, while the single open bp's entropy gain cannot balance the required enthalpy. This contrasts the opening of further bps, for which the entropy gain almost balances the enthalpy cost. The co-

operativity factor σ_0 helps stabilize the coalesced DNA stretch against reclosure, as the combined free energy of the two individual bubbles carries a factor σ_0^2 while the coalesced bubble has only a factor σ_0.

The full partition function is

$$\mathcal{Z}(X,Y) = \mathcal{Z}_L(X)\mathcal{Z}_R(Y). \tag{3}$$

It defines the free energy landscape $\mathcal{F}(X,Y) = -\beta^{-1}\log[\mathcal{Z}(X,Y)]$, in which the random motion of the zipper forks takes place, as the gradient of $\mathcal{F}$ with respect to the coordinates X and Y defines the local driving forces experienced by the two zipper forks.

Below the melting temperature of the barrier T_b, the barrier will on average be driven toward closure, while above T_b it will tend to denature completely. The effect of thermal fluctuations is to introduce a random-walk-type dynamics of the position of the two zipper forks. Eventually, full denaturation of the clamp may be reached even below the melting temperature T_b. Once the two bubbles coalesce, the loop initiation (cooperativity) factor σ_0 is released, and the coalesced state becomes stabilized against closure.

Dynamically we quantify the random motion of the two zipper forks due to thermal fluctuations as follows. To zip close an already opened bp, we assume that this process is mainly governed by diffusion-limited encounter of the two separated bases and subsequent bond formation. In contrast, to unzip a still closed bp, the free energy barrier embodied in the Boltzmann factor u has to be overcome. For the left zipper fork we define $t_L^+(X,Y)$ which is the transfer coefficient for the process $X \to X+1$, corresponding to clamp size decrease, and $t_L^-(X,Y)$ the transfer coefficient for the process $X \to X-1$ (clamp size increase). For the right zipper fork we similarly introduce $t_R^+(X,Y)$ for the process $Y \to Y+1$ (clamp size increase) and $t_R^-(X,Y)$ for the process $Y \to Y-1$ (clamp size decrease). Due to the end clamping we require that $X \geq -N_L$ and $Y \leq N+N_R$, which amounts to introducing reflecting boundary conditions[53]

$$t_L^-(X=-N_L,Y) = 0 \tag{4}$$

and

$$t_R^+(X,Y=N+N_R) = 0. \tag{5}$$

Once the clamp has vanished, we assume that the clamp will not be able to reform for a long time, and we impose the absorbing conditions

$$t_L^-(X,X) = t_R^+(Y,Y) = 0. \tag{6}$$

Dynamically, this is connected to the time it takes the long stretch of single strand to re-establish a bp in the clamp region (diffusion limit). In terms of the free energy the suppression of clamp reformation is due to the release of the free energy ΔG_{bubble} corresponding to the cooperativity factor σ_0 on bubble coalescence (it would cost the additional factor σ_0 to reintroduce two single-strand/double-strand boundaries).

Knowledge of the transfer coefficients together with the boundary conditions above completely determines the dynamics, and we proceed by giving explicit expressions for the transfer coefficients in terms of the physical parameters of the problem. For the zipping rates we choose

$$t_L^-(X,Y) = \tfrac{1}{2}\mathcal{K}(X+N_L) \tag{7}$$

for the left fork, and identically for the right fork

$$t_R^+(X,Y) = \tfrac{1}{2}\mathcal{K}(N_R+N-Y). \tag{8}$$

We defined above a bubble-size-dependent rate coefficient

$$\mathcal{K}(q) = kq^{-\mu}, \tag{9}$$

with q being the number of broken bps in the bubble, where we have, as in previous studies, introduced the hook exponent μ, related to the fact that during the zipping process not only the bp at the zipper fork is moved but also part of the single strand is dragged or pushed along. This additional effect may be included using similar arguments as in Refs. 21, 41, and 54: To zip close a bp, the two single strands making up the bubble have to be pulled closer toward the zipper fork. The adjustment of pulling propagates along the contour of the chain until the closest bend (inflection) is reached, a distance that scales as the gyration radius, i.e., $\simeq q^{\nu}$. Having in mind Rouse-type dynamics, this would slow down the unzipping rates by the factor $q^{-\nu}$. Hydrodynamic interactions may change the exponent and we here take the transfer coefficients above proportional to $q^{-\mu}$, with μ to be determined by more detailed microscopic investigations. The rate constant k appearing in Eq. (9) is the rate constant for pure bp zipping without factors due to the coupling along the chain, i.e., the hook exponent. The factor 1/2 introduced above in Eqs. (7) and (8) is merely for convenience to be consistent with the nomenclature of previous approaches.[21,41] Apart from the hook effect, we thus assign a factor $k/2$ for the zipping at each of the forks.

As the DNA construct is embedded in a thermal bath, we require the zipping rates to fulfill the detailed balance conditions

$$t_L^+(X-1,Y)\mathcal{Z}(X-1,Y) = t_L^-(X)\mathcal{Z}(X,Y) \tag{10}$$

and

$$t_R^-(X,Y+1)\mathcal{Z}(X,Y+1) = t_R^+(X,Y)\mathcal{Z}(X,Y). \tag{11}$$

These conditions guarantee the relaxation to the thermodynamic equilibrium. For the left zipping rates this is fulfilled for

$$t_L^+(X,Y) = \tfrac{1}{2}\mathcal{K}(q_L+1)u(X+1)\{(q_L+1)/(q_L+2)\}^c. \tag{12}$$

Here $q_L = X+N_L$ is the length of the left bubble, and again $[(\text{bubble length})+1]^{-c}$ is the correction for the entropy loss of a closed polymer loop, with c being the loop exponent. For bubble-size increase we thus take the transfer coefficients to be proportional to the Arrhenius factor $u(X+1)$, multiplied by a loop correction factor. Note that for $q_L \to \infty$ the loop correction factor tends to 1, which we will exploit later. For the right fork we similarly find

$$t_R^-(X,Y) = \tfrac{1}{2}\mathcal{K}(q_R+1)u(Y-1)\{(q_R+1)/(q_R+2)\}^c, \tag{13}$$

where $q_R = N+N_R-Y$ is the length of the right bubble. We point out that Eqs. (7), (8), (12), and (13) are not unique in satisfying the detailed balance conditions, Eqs. (10) and (11). However, different choices correspond to redefinitions of the

time unit $1/k$ which is the free parameter in our master equation approach and needs to be fixed from fit to experiment.

From the transition rates a master equation can be constructed for the conditional probability $P(X,Y;t|X_0,Y_0)$ with X_0,Y_0 being the initial positions of the zipper forks. This master equation can be solved numerically, details of which are introduced in Sec. VI, and physical quantities such as the mean-first-passage time density can be calculated. However, based on four assumptions concerning the transition rates it is possible to derive a continuous Fokker–Planck equation approximating the full master equation description; this is done in Sec. III. From the Fokker–Planck approach we then derive numerical results for the coalescence time density and both numerical and analytic expressions for the mean coalescence time and the probability density for the coalescence position in Secs. IV and V. Sections III–V provide details and extensions of our previous short work.[55]

III. THE FOKKER–PLANCK APPROXIMATION TO THE MASTER EQUATION

In this section we derive a Fokker–Planck approximation to the master equation based on the following assumptions:

(i) the temperature T is so high compared to T_s that the bps in the soft zones remain unzipped at all times, i.e., there are effectively reflecting boundary conditions at the interfaces between the soft zones and the barrier region;
(ii) the soft zones are sufficiently long such that the influence of the loop factors can be neglected;
(iii) similarly, the influence of the hook factors becomes sufficiently small; and
(iv) finally, the number of bps in the barrier region is much bigger than 1, i.e., $N \gg 1$, which allows taking the continuum limit (see below).

Under assumptions (i)–(iii) the full partition function for the two bubbles and the partially denatured barrier region becomes [see Eq. (3)]

$$\mathcal{Z} = \sigma_0^2 u_s^{N_L+N_R} u_b^{X+N-Y}, \tag{14}$$

where in this section $X,Y \in \{0,1,2,\ldots,N\}$ due to (i) with $X \leq Y$. Notice that X is the number of barrier bps already broken from the left end of the barrier, and $N-Y$ counts the broken barrier bps from the right end. The free energy is given as

$$\mathcal{F} = -2k_BT \log \sigma_0 - (N_L+N_R)\Delta G_s + (X+N-Y)\Delta G_b. \tag{15}$$

In the continuum limit [assumption (iv)] we introduce dimensionless coordinates $x=X/N$ and $y=Y/N$, $x,y \in [0,1]$. The gradient of $\mathcal{F}$ with respect to the coordinates x and y defines the local force experienced by the two zipper forks, namely,

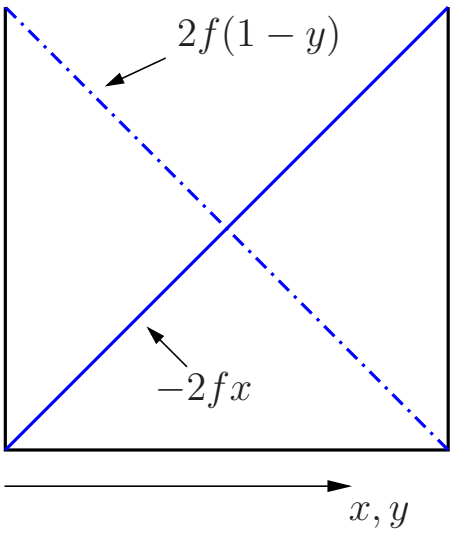

FIG. 2. (Color online) Plot of the linear potentials experienced by the respective bubble interfaces in the case $T<T_b$ ($f<0$) in terms of the dimensionless quantities x, y, and f (see text for details).

$$F_X = -d\mathcal{F}/dx = -N\Delta G_b, \quad F_Y = -d\mathcal{F}/dy = N\Delta G_b = -F_X, \tag{16}$$

and we immediately see that the zipper forks X and Y are driven by opposite, constant forces as sketched in Fig. 2.

Using simplifications (ii) and (iii) stated above, the modified continuum rates (denoted by r to distinguish them from the notation introduced in the discussion of the discrete case) for closing a bp at the left fork at position x or at the right fork at position y become [see Eqs. (7), (8), (12), and (13)]

$$r_L^-(x,y) = r_R^+(x,y) = k/2 \tag{17}$$

and

$$r_L^+(x,y) = r_R^-(x,y) = u_b k/2, \tag{18}$$

such that the zipping open of a bp requires crossing the barrier ΔG_b. The boundary conditions

$$r_L^-(0,y) = r_R^+(x,1) = r_L^+(x,x) = r_R^-(y,y) = 0 \tag{19}$$

guarantee that bps cannot close beyond the barrier region, and that the process ends when the two zipper forks coalesce.

Define by $P(x,y;\tau)$ the probability density that the left and right zipper forks are located at x and y, respectively, at some given time τ. The time evolution of $P(x,y;\tau)$ is then given in terms of the master equation[56,57]

$$\begin{aligned}\frac{\partial}{\partial\tau}P(x,y;\tau) &= r_L^+(x-1/N,y)P(x-1/N,y;\tau) \\ &+ r_L^-(x+1/N,y)P(x+1/N,y;\tau) \\ &- [r_L^-(x,y)+r_L^+(x,y)]P(x,y;\tau) \\ &+ r_R^-(x,y+1/N)P(x,y+1/N;\tau) \\ &+ r_R^+(x,y-1/N)P(x,y-1/N;\tau) \\ &- [r_R^-(x,y)+r_R^+(x,y)]P(x,y;\tau).\end{aligned} \tag{20}$$

Following the standard derivation[57,58] we Taylor expand the

above master equation keeping the first two orders only. For instance, for the first term on the right hand side of Eq. (20), we obtain the Taylor expansion

$$r_L^+(x-1/N,y)P(x-1/N,y;\tau) \approx r_L^+(x,y)P(x,y;\tau) - \frac{1}{N}\frac{\partial}{\partial x}r_L^+(x,y)P(x,y;\tau) + \frac{1}{2N^2}\frac{\partial^2}{\partial x^2}r_L^+(x,y)P(x,y;\tau). \tag{21}$$

This is the only consistent expansion of finite order according to the Pawula–Maricinkiewicz theorem.[56,59,60] Alternatively the full Kramers–Moyal expansion needs to be taken along. With analogous expansions for the other terms and after some rearrangement, we find the bivariate Fokker–Planck equation[56]

$$\frac{\partial}{\partial\tau}P(x,y;\tau|x_0,y_0) = F\left(\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right)P(x,y;\tau|x_0,y_0) + D\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)P(x,y;\tau|x_0,y_0), \tag{22}$$

where, instead of the probability density $P(x,y;\tau)$, we use explicitly the notation $P(x,y;\tau|x_0,y_0)$ including the initial conditions x_0 and y_0. In Eq. (22), the force F and diffusion constant D are defined by

$$F \equiv \frac{k(u_b-1)}{2N} \tag{23}$$

and

$$D \equiv \frac{k(u_b+1)}{4N^2}. \tag{24}$$

Equation (22) is completed by specifying the initial and boundary conditions. As initial condition, we choose the sharp δ form

$$P(x,y;0|x_0,y_0) = \delta(x-x_0)\,\delta(y-y_0), \tag{25}$$

with $x_0<y_0$, and due to the initial condition Eq. (25) the joint probability density $P(x,y;\tau|x_0,y_0)$ is actually Green's function of Eq. (22). The condition that the two bubbles in the soft zones are always open is guaranteed by the reflecting boundary conditions (here, we define $2f\equiv F/D$)

$$\left(\frac{\partial}{\partial x}-2f\right)P(x,y;\tau|x_0,y_0)|_{x=0}=0, \qquad \left(\frac{\partial}{\partial y}+2f\right)P(x,y;\tau|x_0,y_0)|_{y=1}=0 \tag{26}$$

at the edges of the line segment [0,1]: Once a zipper fork reaches either edge, the only possible direction to move is to restart unzipping the barrier. Moreover, we specify the viciousness condition[61]

$$P(x,x;\tau|x_0,y_0)=0, \tag{27}$$

according to which the two zipper forks cannot be at the same point: the two forks annihilate and the bubbles coalesce. This last condition ensures the continuous character of the probability density P. For completeness, we actually need to specify a second set of boundary conditions. However, due to the viciousness condition (27), we can choose this boundary condition *ad libitum*; a clever choice will turn out to be

$$\frac{\partial}{\partial x}P(x,y;\tau|x_0,y_0)|_{x=1}=0, \qquad \frac{\partial}{\partial y}P(x,y;\tau|x_0,y_0)|_{y=0}=0. \tag{28}$$

Such a choice is possible because the zipper forks never reach these two points.

IV. SOLUTION OF THE VICIOUS WALKER PROBLEM

A. Transformation of the Fokker–Planck equation

To obtain the solution of the Fokker–Planck equation (22), it is convenient to notice that after a redefinition of time unit $t=D\tau$ the problem depends on a single dimensionless parameter

$$f=\frac{F}{2D}=N\frac{u_b-1}{u_b+1}. \tag{29}$$

It is important that this parameter depends on the length of the barrier N and the Boltzmann factor u_b for opening the barrier bps but not on the kinetic constant k. Thus, apart from an overall prefactor fixing the time unit, the solution depends solely on the structural properties of the physical system under study.

Let us summarize the rephrased problem in terms of f for completeness,

$$\left[-\frac{\partial}{\partial t}+\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}-2f\frac{\partial}{\partial x}+2f\frac{\partial}{\partial y}\right]P(x,y;t|x_0,y_0)=0, \tag{30a}$$

with boundary conditions

$$\left(\frac{\partial}{\partial x}-2f\right)P(x,y;t|x_0,y_0)|_{x=0}=0, \tag{30b}$$

$$\left(\frac{\partial}{\partial y}+2f\right)P(x,y;t|x_0,y_0)|_{y=1}=0, \tag{30c}$$

$$\frac{\partial}{\partial x}P(x,y;t|x_0,y_0)|_{x=1}=0, \tag{30d}$$

$$\frac{\partial}{\partial y}P(x,y;t|x_0,y_0)|_{y=0}=0, \tag{30e}$$

the viciousness condition

$$P(x,x;t|x_0,y_0) = 0, \tag{30f}$$

and the initial condition

$$P(x,y;0|x_0,y_0) = \delta(x-x_0)\,\delta(y-y_0) \quad \text{with } x_0 < y_0. \tag{30g}$$

To proceed in the solution, let us first introduce the Fokker–Planck operators

$$\mathbb{L}^{+}_{\mathrm{FP}}(x) \equiv \frac{\partial^2}{\partial x^2} - 2f\frac{\partial}{\partial x}, \tag{31a}$$

$$\mathbb{L}^{-}_{\mathrm{FP}}(y) \equiv \frac{\partial^2}{\partial y^2} + 2f\frac{\partial}{\partial y}, \tag{31b}$$

so that the Fokker–Planck equation (30) can be recast into the following form:

$$\frac{\partial}{\partial t}P(x,y;t|x_0,y_0) = [\mathbb{L}^{+}_{\mathrm{FP}}(x) + \mathbb{L}^{-}_{\mathrm{FP}}(y)]P(x,y;t|x_0,y_0). \tag{32}$$

The operators $\mathbb{L}^{+}_{\mathrm{FP}}(x)$ and $\mathbb{L}^{-}_{\mathrm{FP}}(y)$ can now be used to transform our Fokker–Planck equation following the procedure outlined in Ref. 56, Chap. 5.4. Let us now define the Hermitian operator

$$\mathbb{L}(x,x_0) \equiv e^{-f(x-x_0)}\mathbb{L}^{+}_{\mathrm{FP}}(x)e^{f(x-x_0)} = \frac{\partial^2}{\partial x^2} - f^2. \tag{33}$$

Here the last equality sign can be shown by applying the operator to a test function. The operator $\mathbb{L}(x,x_0)$ corresponds to a Hamilton operator of a Schrödinger equation with imaginary time $-i\hbar t$, mass $m=\hbar^2/2$, and constant potential f^2. The idea is that it is (in general) easier to solve the time-dependent Schrödinger equation than the Fokker–Planck equation because the first-order derivative has been eliminated. The relation between the solutions of the original and the transformed Fokker–Planck equations are found after a few lines of algebra, and we obtain the following result: If $\tilde{p}^L(x;t|x_0)$ is a solution of

$$\frac{\partial}{\partial t}\tilde{p}^L(x;t|x_0) = \mathbb{L}(x,x_0)\tilde{p}^L(x;t|x_0), \tag{34}$$

then the density

$$p^L(x;t|x_0) = \exp[f(x-x_0)]\tilde{p}^L(x;t|x_0) \tag{35}$$

solves the original equation

$$\frac{\partial}{\partial t}p^L(x;t|x_0) = \mathbb{L}^{+}_{\mathrm{FP}}(x)p^L(x;t|x_0). \tag{36}$$

The boundary conditions in x, Eqs. (30b) and (30d), are transformed to

$$\left(\frac{\partial}{\partial x} - f\right)\tilde{p}^L(x;t|x_0)\big|_{x=0} = 0, \tag{37a}$$

$$\left(\frac{\partial}{\partial x} + f\right)\tilde{p}^L(x;t|x_0)\big|_{x=1} = 0, \tag{37b}$$

while the initial condition remains unchanged, $\tilde{p}^L(x;t=0|x_0) = \exp[-f(x-x_0)]p^L(x;t=0|x_0) = \delta(x-x_0)$. In other words, with Eqs. (33), (37a), and (37b) we have indeed transformed the original Fokker–Planck operator $\mathbb{L}^{+}_{\mathrm{FP}}(x)$ into Hermitian form.

Noticing that

$$\mathbb{L}(y,y_0) = \frac{\partial^2}{\partial y^2} - f^2 = e^{f(y-y_0)}\mathbb{L}^{-}_{\mathrm{FP}}(y)e^{-f(y-y_0)}, \tag{38}$$

the same procedure can be carried out for the y coordinate, where the different sign in front of f in Eq. (30a) causes that, if $\tilde{p}^R(y;t|y_0)$ is a solution to

$$\frac{\partial}{\partial t}\tilde{p}^R(y;t|y_0) = \mathbb{L}(y,y_0)\tilde{p}^R(y;t|y_0), \tag{39}$$

then

$$p^R(y;t|y_0) = \exp[-f(y-y_0)]\tilde{p}^R(y;t|y_0) \tag{40}$$

satisfies

$$\frac{\partial}{\partial t}p^R(y;t|y_0) = \mathbb{L}^{-}_{\mathrm{FP}}(y)p^R(y;t|y_0), \tag{41}$$

and the boundary conditions are

$$\left(\frac{\partial}{\partial y} + f\right)\tilde{p}^R(y;t|y_0)\big|_{y=1} = 0, \tag{42a}$$

$$\left(\frac{\partial}{\partial y} - f\right)\tilde{p}^R(y;t|y_0)\big|_{y=0} = 0, \tag{42b}$$

together with $\tilde{p}^R(y;t=0|y_0) = \delta(y-y_0)$.

We now see why Eqs. (30d) and (30e) are clever choices for the additional boundary conditions on $P(x,y;t|x_0,y_0)$, together with the similarity transformations (35) and (40): reflecting the uneven symmetry of the problem with respect to the original coordinates, *all* equations defining the functions $\tilde{p}^L(x;t|x_0)$ and $\tilde{p}^R(x;t|x_0)$ are identical, and thus are the functions themselves,

$$\tilde{p}^L(x;t|x_0) = \tilde{p}^R(x;t|x_0) = \tilde{p}(x;t|x_0). \tag{43}$$

B. Solution of the transformed Fokker–Planck equation

After the individual similarity transformations for the left [Eq. (35)] and right [Eq. (40)] walkers are performed, we arrive at the following time-dependent Schrödinger equation:

$$\frac{\partial}{\partial t}\tilde{P}(x,y;t|x_0,y_0) = [\mathbb{L}(x,x_0) + \mathbb{L}(y,y_0)]\tilde{P}(x,y;t|x_0,y_0), \tag{44}$$

with imaginary time; here,

$$\tilde{P}(x,y;t|x_0,y_0) = e^{-f(x-x_0)+f(y-y_0)}P(x,y;t|x_0,y_0), \tag{45}$$

which implies the same viciousness condition $\tilde{P}(x,x;t|x_0,y_0)=0$ as in the original formulation. Now, however, we effectively have two identical vicious walkers moving in a common (constant) potential for which the solution is well known.[50,51] It is given by the antisymmetric product of the single-walker solutions, namely,

$$\tilde{P}(x,y;t|x_0,y_0) = \tilde{p}(x;t|x_0)\tilde{p}(y;t|y_0) - \tilde{p}(y;t|x_0)\tilde{p}(x;t|y_0). \tag{46}$$

Note that this form of the solution is analogous to constructing the solution of an absorbing boundary value problem for a single diffusor under a constant drift according to the method of images.[62]

The backward transformation of solution (46) by inverting Eq. (45) finally produces

$$P(x,y;t|x_0,y_0) = e^{f(x-x_0)-f(y-y_0)}[\tilde{p}(x;t|x_0)\tilde{p}(y;t|y_0) - \tilde{p}(y;t|x_0)\tilde{p}(x;t|y_0)], \tag{47}$$

and by construction this is a solution of Eq. (30a), satisfying the boundary conditions Eqs. (30b)–(30e), as well as the viciousness condition $P(x,x;t|x_0,y_0)=0$. It remains to check that this solution also satisfies the initial condition (30) and, thus, fully solves the studied problem. For $t=0$ we obtain

$$\begin{aligned} &P(x,y;t=0|x_0,y_0) \\ &= e^{f(x-x_0)-f(y-y_0)}[\delta(x-x_0)\delta(y-y_0) - \delta(y-x_0)\delta(x-y_0)] \\ &= \delta(x-x_0)\delta(y-y_0) - e^{2f(y_0-x_0)}\delta(y-x_0)\delta(x-y_0) \\ &= \delta(x-x_0)\delta(y-y_0), \end{aligned} \tag{48}$$

valid for $x_0<y_0$ and $x<y$. The last equality follows because of the choice of the initial condition $x_0<y_0$ and the viciousness of the process, i.e., the walkers can never pass each other, which also implies $x<y$ for any realizable configuration at any time t. Thus, the arguments of the δ functions in the second term can never vanish simultaneously and therefore this term is effectively equal to zero. Thus Eq. (47) is the solution to our problem.

C. Calculation of $\tilde{p}(x;t|x_0)$ and its spectral resolution

To find the solution of the full two-walker problem (47) we need to calculate $\tilde{p}(x;t|x_0)$, which solves the Schrödinger equation

$$\frac{\partial}{\partial t}\tilde{p}(x;t|x_0) = \left[\frac{\partial^2}{\partial x^2} - f^2\right]\tilde{p}(x;t|x_0) \tag{49}$$

and satisfies the boundary conditions Eqs. (37a) and (37b), as well as the initial condition $\tilde{p}(x;t=0|x_0)=\delta(x-x_0)$. Unfortunately, there is no explicit solution of this equation in the time domain. It can be found, however, in the Laplace picture which is done in detail in Appendix A. Here we only summarize the final result,

$$\tilde{p}(x;z|x_0) = \frac{1}{2k(\kappa^2 e^{2k}-1)}[e^{k|x-x_0|} + \kappa^2 e^{2k}e^{-k|x-x_0|} + \kappa e^{(x+x_0)k} + \kappa e^{2k}e^{-(x+x_0)k}], \tag{50}$$

with z-dependent

$$k \equiv k(z) = \sqrt{z+f^2} \tag{51}$$

and

$$\kappa \equiv \kappa(z) = \frac{k(z)+f}{k(z)-f}. \tag{52}$$

The corresponding behavior in the time domain is found by the inverse Laplace transform

$$\begin{aligned} \tilde{p}(x;t|x_0) &= \int_{\epsilon-i\infty}^{\epsilon+i\infty}\frac{dz}{2\pi i}e^{zt}\tilde{p}(x;z|x_0) \\ &= \sum_{n=0}^{\infty} e^{\lambda_n t}\psi_n(x)\psi_n^*(x_0), \end{aligned} \tag{53}$$

with $\epsilon \geq 0$ large enough such that all singularities of $\tilde{p}(x;z|x_0)$ lie in the half-plane $\mathrm{Re}\, z<\epsilon$. We used the formal eigenmode expansion in the second line of the above equation which exists as a spectral resolution of the Hermitian operator defined by Eq. (49) and the pertinent boundary conditions (37a) and (37b). The eigenvalues λ_n are real numbers since the operator is Hermitian but not necessarily nonpositive like in the case of standard Fokker–Planck operators in one dimension. The reason is that $\tilde{p}(x;t|x_0)$ is an auxiliary mathematical quantity without any direct physical meaning and, thus, can in principle grow exponentially over time, i.e., some of the λ_n might be positive. The physical quantity which is not allowed to grow indefinitely is the two-walker Green's function (47). This is useful to keep in mind when studying the spectral resolutions of (50) and (47) in more detail in the following.

The spectrum of the operator $\partial^2/\partial x^2 - f^2$ from Eq. (49), together with the boundary conditions, can be found either directly from the defining equations or by finding the poles of the one-walker Green's function (50). Indeed, the secular equation obtained by either method is equivalent to the denominator in Eq. (50) being zero, i.e., $\kappa^2(\lambda)=e^{-2k(\lambda)}$. This leads to transcendent equations for real λ in the respective ranges

$$(\lambda+2f^2)\sinh\sqrt{\lambda+f^2} + 2f\sqrt{\lambda+f^2}\cosh\sqrt{\lambda+f^2} = 0 \quad \text{for } \lambda \geq -f^2, \tag{54a}$$

$$(\lambda+2f^2)\sin\sqrt{|\lambda|-f^2} + 2f\sqrt{|\lambda|-f^2}\cos\sqrt{|\lambda|-f^2} = 0 \quad \text{for } \lambda \leq -f^2. \tag{54b}$$

One looks for solutions of these equations in their ranges of validity. Thus, with the help of the standard graphical analysis, it turns out that Eq. (54) has at least one solution only for $f<0$; this solution is positive, i.e., $\lambda_0>0$. When $f<-2$ a further solution with $-f^2<\lambda_1<0$ solves Eq. (54a). There are no more options for Eq. (54a). On the other hand, Eq. (54b) always has an infinite number of solutions. For $f>0$ these solutions are bounded by $-f^2-(n+1)^2\pi^2 \leq \lambda_n \leq -f^2-n^2\pi^2$ $(n=0,1,2,\ldots)$. For $0>f>-2$ the first eigenvalue λ_0 satisfies Eq. (54b) while the remaining eigenvalues (λ_n for $n=1,2,\ldots$) stem from Eq. (54b) and are still bounded by $-f^2-n^2\pi^2 \leq \lambda_n \leq -f^2-(n-1)^2\pi^2$. Finally, for $f<-2$ the first two eigenvalues λ_0 and λ_1 are determined by Eq. (54a) and the rest stems from Eq. (54b) being bounded by $-f^2-n^2\pi^2 \leq \lambda_n \leq -f^2-(n-1)^2\pi^2$ for $n=2,3,\ldots$.

There are two special values of the force $f=0,-2$ where the spectrum appears to change its analytic structure. First, the $f=0$ case corresponds to the problem of two vicious walkers freely diffusing in an impenetrable well, i.e., in a common potential. For this case, we do not need to use our trick as it is solved already by a previous study:[51] The single-walker spectrum reads $\lambda_n=(n\pi)^2$ for $n=0,1,2,\ldots$. The second case, $f=-2$, does not appear to be in any way particular physically. Actually, neither of the two cases have any exceptional physical properties—all the physical quantities change smoothly across these two points when changing f. The apparent singularities occur only in the auxiliary quantities. Technically, these two cases are the only ones where the trivial solution $\lambda=-f^2$ of Eq. (54) corresponds to a nontrivial, i.e., nonzero, solution for the eigenfunction. Mathematically, this is reflected by the formal failure when $k(z)=0$ of the method used in Appendix A leading to Eq. (50). In such cases, the two fundamental solutions $e^{\pm kx}$ of Eq. (A3) become identical and equal to a constant. The second independent solution is then linear in x according to the elementary theory of linear differential equations. Thus, the two special cases are not directly covered by the general solution presented in Appendix A and all formulas stemming from it. We do not give explicit solutions for those two singular cases since all quantities of interest can be obtained from the general formulas by taking the appropriate limit $f\to 0$ or -2.

Now, if we insert expansion (53) into Eq. (47) we see that the spectrum of that equation is given by all pairwise sums of *different* one-walker eigenvalues, i.e., $\Lambda_k=\lambda_i+\lambda_j (i\neq j)$ [the sums of identical one-walker eigenvalues have zero weight due to the antisymmetrization in Eq. (47)]. Therefore, the first two one-walker solutions λ_0 and λ_1 determine the long-time asymptotics of the two-walker problem. For large negative $f\ll -1$, corresponding to a large barrier case, the asymptotic behavior dominates the whole solution and, thus, the combination $\Lambda_0\equiv\lambda_0+\lambda_1$ effectively determines the interesting quantities such as the mean coalescence time or the probability distribution of the coalescence position. This allows us to get full analytic results in the limit $f\ll -1$, as explicitly derived in Sec. V B. In this case both eigenvalues $\lambda_{0,1}$ are solutions of Eq. (54a), and they are exponentially small in $|f|$. They can be found to leading order from the second order expansion in λ of Eq. (54a), yielding $\lambda_{0,1}\simeq\pm 4f^2e^{-|f|}$. To obtain Λ_0 we need to increase the accuracy by expanding the equation up to the third order in λ and we find $\Lambda_0=\lambda_0+\lambda_1\simeq -16f^2(|f|-1)e^{-2|f|}$ which is negative, as it should be since $P(x,y;t|x_0,y_0)$ cannot grow exponentially, P being a physical quantity. So, despite the existence of a positive one-walker eigenvalue λ_0, the physically relevant combination $\lambda_0+\lambda_1<0$ ensures meaningful results.

V. BUBBLE COALESCENCE TIME AND POSITION FROM THE SOLUTION OF THE FOKKER–PLANCK EQUATION

A. General treatment

From the solution of the Fokker–Planck equation, we now calculate in general the quantities of interest, namely, the characteristic coalescence time of the two random walking zipper forks and the probability distribution of the coalescence position. In the time domain this is done by considering the conservation of probability in the form

$$\Pi(t|x_0,y_0)+\int_0^1 dy\int_0^y dx P(x,y;t|x_0,y_0)=1, \tag{55}$$

which is actually a defining equation for the probability $\Pi(t|x_0,y_0)$ that the walkers have met before time t. Consequently, the second term represents the probability of having two separate bubbles at time t (survival probability). Note the range of integration restricting $x\in[0,y)$. The probability density associated with $\Pi(t|x_0,y_0)$ is

$$\begin{aligned}\pi(t|x_0,y_0)&\equiv\frac{d}{dt}\Pi(t|x_0,y_0)\\&=-\int_0^1 dy\int_0^y dx\frac{\partial}{\partial t}P(x,y;t|x_0,y_0),\end{aligned} \tag{56}$$

which, after invoking Eq. (30) and rearranging the order of the integrals, yields

$$\begin{aligned}\pi(t|x_0,y_0)&=-\int_0^1 dy\int_0^y dx\left[\frac{\partial^2}{\partial x^2}-2f\frac{\partial}{\partial x}\right]P(x,y;t|x_0,y_0)-\int_0^1 dx\int_x^1 dy\left[\frac{\partial^2}{\partial y^2}+2f\frac{\partial}{\partial y}\right]P(x,y;t|x_0,y_0)\\&=-\int_0^1 dy\left[\left\{\frac{\partial}{\partial x}-2f\right\}P(x,y;t|x_0,y_0)\right]_0^y-\int_0^1 dx\left[\left\{\frac{\partial}{\partial y}+2f\right\}P(x,y;t|x_0,y_0)\right]_x^1\\&=-\int_0^1 dy\left\{\frac{\partial}{\partial x}-2f\right\}P(x,y;t|x_0,y_0)\Big|_{x=y}+\int_0^1 dx\left\{\frac{\partial}{\partial y}+2f\right\}P(x,y;t|x_0,y_0)\Big|_{y=x}\\&=\int_0^1 dx\left\{\frac{\partial}{\partial y}-\frac{\partial}{\partial x}+4f\right\}P(x,y;t|x_0,y_0)\Big|_{y=x}\\&=\int_0^1 dx\left\{\frac{\partial}{\partial y}-\frac{\partial}{\partial x}\right\}P(x,y;t|x_0,y_0)\Big|_{y=x},\end{aligned} \tag{57}$$

where the boundary conditions imposed on $P(x,y;t|x_0,y_0)$ and the viciousness condition have been used in subsequent manipulations. The quantity $\pi(t|x_0,y_0)$ can be interpreted as the probability current flowing into the absorbing boundary, here given by the line $x=y$ (compare with an analogous discussion for a one-dimensional case in Ref. 63).

From the last relation, we define the quantity $\varrho(x;t|x_0,y_0)$, the coalescence-position-resolved probability density for the coalescence time,

$$\pi(t|x_0,y_0) = \int_0^1 dx \varrho(x;t|x_0,y_0), \tag{58}$$

such that

$$\varrho(x;t|x_0,y_0) = \left\{ \frac{\partial}{\partial y} - \frac{\partial}{\partial x} \right\} P(x,y;t|x_0,y_0)|_{y=x}. \tag{59}$$

We are mainly interested in either the mean coalescence time $\tau(x_0,y_0)=\int_0^\infty dt t \pi(t|x_0,y_0)$ or the probability density of the coalescence position, $\rho(x|x_0,y_0)=\int_0^\infty dt \varrho(x;t|x_0,y_0)$. These are quantities integrated over time and, thus, they may be determined from the solution in the Laplace domain without explicit knowledge of $P(x,y;t|x_0,y_0)$ in the time domain.

We now rewrite Eq. (47) in terms of the inverse Laplace transforms as follows:

$$\begin{aligned} P(x,y;t|x_0,y_0) = & \int_{\epsilon-i\infty}^{\epsilon+i\infty} \frac{dz_1}{2\pi i} e^{z_1 t} \int_{\epsilon-i\infty}^{\epsilon+i\infty} \frac{dz_2}{2\pi i} e^{z_2 t} e^{f(x-y-x_0+y_0)} \\ & \times [\tilde{p}(x;z_1|x_0)\tilde{p}(y;z_2|y_0) \\ & - \tilde{p}(y;z_1|x_0)\tilde{p}(x;z_2|y_0)], \end{aligned} \tag{60}$$

where $\tilde{p}(x;z|x_0)$ is given in Eq. (50). In case that $\tilde{p}(x;z|x_0)$ has no poles in the half-plane Re $z>0$ (corresponding to $f>0$, $\epsilon=0$), we can directly use the substitutions $z_1=z/2+i\omega$ and $z_2=z/2-i\omega$, $z=z_1+z_2$, and $\omega=(z_1-z_2)/(2i)$ to obtain a single inverse Laplace transform for $P(x,y;t|x_0,y_0)$, namely,

$$\begin{aligned} P(x,y;t|x_0,y_0) = & \int_{-i\infty}^{i\infty} \frac{dz}{2\pi i} e^{zt} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{f(x-y-x_0+y_0)} \\ & \times [\tilde{p}(x;z/2+i\omega|x_0)\tilde{p}(y;z/2-i\omega|y_0) \\ & - \tilde{p}(y;z/2+i\omega|x_0)\tilde{p}(x;z/2-i\omega|y_0)]. \end{aligned} \tag{61}$$

Thus, we can identify the Laplace transform $\tilde{P}(x,y;z|x_0,y_0)$ of $P(x,y;t|x_0,y_0)$ for $f>0$ as

$$\begin{aligned} \tilde{P}(x,y;z|x_0,y_0) = & \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{f(x-y-x_0+y_0)} \\ & \times [\tilde{p}(x;z/2+i\omega|x_0)\tilde{p}(y;z/2-i\omega|y_0) \\ & - \tilde{p}(y;z/2+i\omega|x_0)\tilde{p}(x;z/2-i\omega|y_0)]. \end{aligned} \tag{62}$$

This directly yields, by means of relations (58) and (59), the Laplace transforms $\tilde{\pi}(z|x_0,y_0)$ and $\tilde{\varrho}(x;z|x_0,y_0)$, from which we in turn deduce the (time-averaged) distribution of coalescence positions,

$$\rho(x|x_0,y_0) = \tilde{\varrho}(x;z=0^+|x_0,y_0), \tag{63}$$

and the characteristic (mean) coalescence time

$$\tau(x_0,y_0) = -\left. \frac{d}{dz} \tilde{\pi}(z|x_0,y_0) \right|_{z=0^+}. \tag{64}$$

For $f<0$ the situation is more complicated since $\tilde{p}(x;z|x_0)$ in this case has a pole at $\lambda_0>0$ which prohibits us from just repeating the above reasoning. The Laplace transform (60) only holds for Re $z_{1,2}>\lambda_0>0$ implying Re $z>2\lambda_0>0$ and the analytic continuation down to $z=0$ is not obvious. However, since we know the position of the only pole located in the Re $z>0$ half-plane, we can treat this singularity separately and thus generalize the previous results. Using the Cauchy theorem for complex integrals we move the integration line from $\epsilon+i\omega$ with $\epsilon>\lambda_0>0$ down to the imaginary axis and add the contribution from the (single) singularity at λ_0 in between these two lines. This method eventually leads to an expression for the Laplace transform $P(x,y;z|x_0,y_0)$ in the whole half-plane Re $z>0$ for $f<0$,

$$\begin{aligned} & \tilde{P}(x,y;z|x_0,y_0) \\ & = e^{f(x-y-x_0+y_0)} \left\{ \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} [\tilde{p}(x;z/2+i\omega|x_0)\tilde{p}(y;z/2-i\omega|y_0) \right. \\ & \quad - \tilde{p}(y;z/2+i\omega|x_0)\tilde{p}(x;z/2-i\omega|y_0)] \\ & \quad + \mathrm{Res}_{z_1=\lambda_0} \tilde{p}(x;z_1|x_0)\tilde{p}(y;z-\lambda_0|y_0) \\ & \quad - \mathrm{Res}_{z_1=\lambda_0} \tilde{p}(y;z_1|x_0)\tilde{p}(x;z-\lambda_0|y_0) \\ & \quad + \tilde{p}(x;z-\lambda_0|x_0)\mathrm{Res}_{z_2=\lambda_0} \tilde{p}(y;z_2|y_0) \\ & \quad \left. - \tilde{p}(y;z-\lambda_0|x_0)\mathrm{Res}_{z_2=\lambda_0} \tilde{p}(x;z_2|y_0) \right\}, \end{aligned} \tag{65}$$

which is the sought-for generalization of Eq. (62). We observe that the double-residue term drops out due to the antisymmetrization procedure.

B. Results

Equations (62) and (65) for $f>0$ and $f<0$, respectively, together with the single-walker Green's function (50) and the identities for $\rho(x|x_0,y_0)$ [Eq. (63)] and $\tau(x_0,y_0)$ [Eq. (64)] were implemented in MATHEMATICA and evaluated. The results are shown in Figs. 3 and 4 depicting the coalescence position probability density for several values of f and the mean coalescence time as a function of f, respectively. Both quantities are shown for two different initial conditions $x_0=0$, $y_0=1$ (the two walkers start out right at the boundaries) and $x_0=0.5$, $y_0=0.9$ (a generic initial condition). Further results for the coalescence time probability density $\pi(t|x_0,y_0)$ [Eq. (57)] are presented in later sections.

The results for the coalescence position in Fig. 3 show for both initial conditions a clear crossover from a peaked form of the probability density for large positive force [the case of an almost "free fall" (ff) into the potential well where the boundary conditions have negligible influence on the dynamics, studied in detail in Ref. 31 and discussed below] to

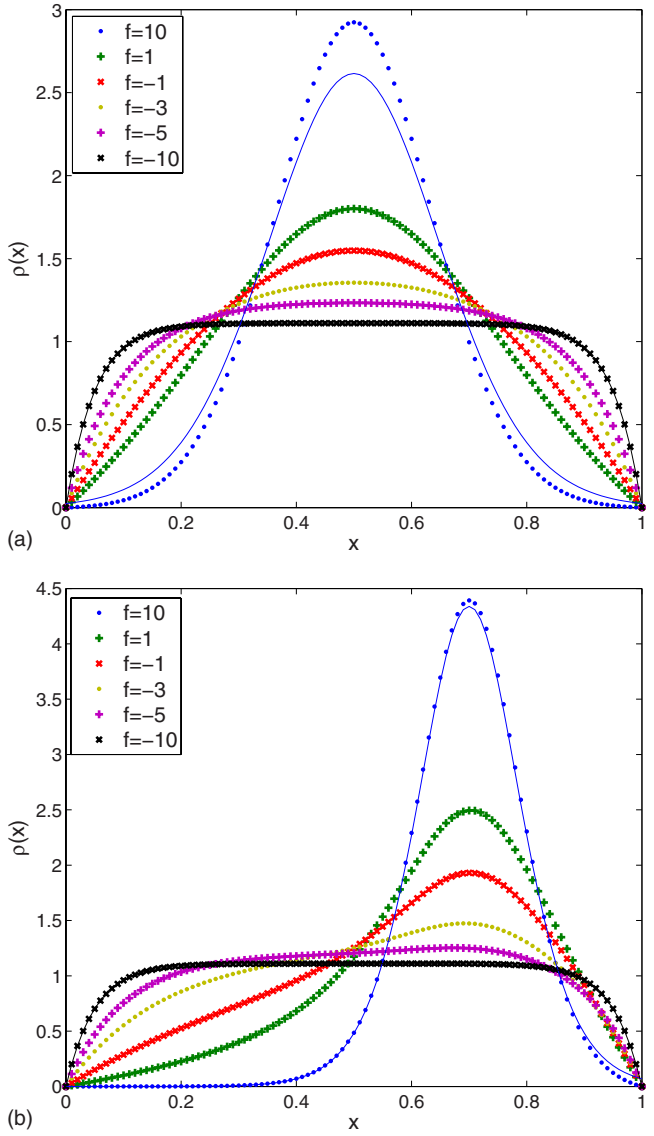

FIG. 3. (Color online) Probability density for the coalescence position $\rho(x|x_0,y_0)$ as a function of the position x for several values of the dimensionless force f. The initial positions of the walkers x_0 and y_0 were (a) $x_0=0$, $y_0=1$ and (b) $x_0=0.5$, $y_0=0.9$. The full lines for cases $f=10$ and $f=-10$ correspond to analytical results in Eqs. (69) and (74), respectively, for given initial conditions.

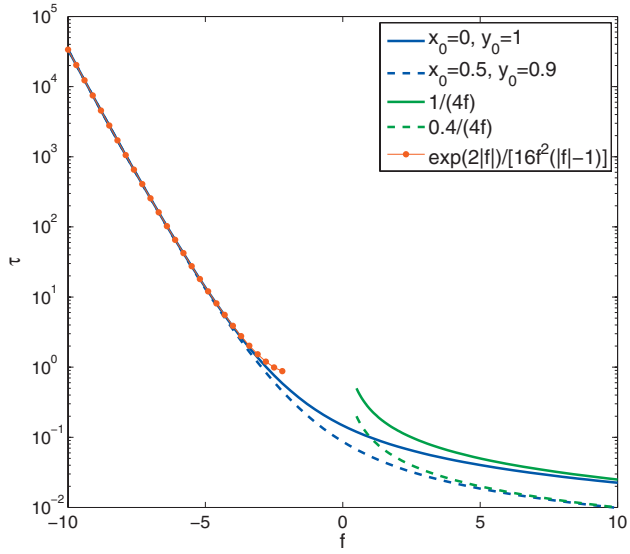

FIG. 4. (Color) The mean coalescence time $\tau(x_0,y_0)$ as a function of the dimensionless force f for two different initial conditions $x_0=0$, $y_0=1$ (full line) and $x_0=0.5$, $y_0=0.9$ (dashed line). The asymptotic analytical results for large positive, i.e., ff case with $\tau=(y_0-x_0)/(4f)$, and negative, i.e., large barrier case of Eq. (75), forces are also shown for comparison.

a very flat probability density in the case of large negative force corresponding to a high barrier. The flatness in the latter case can be understood from a simple Arrhenius-like model, in which the probability of the walker to be at a place x is proportional to a Boltzmann weight, i.e., $\exp[-\beta\phi(x)]$, where $\phi(x)=-\int^x F(x')dx'$ is the free energy corresponding to the force $F(x)=\pm f$. Since we are now dealing with two walkers, the probability of both of them being simultaneously at the coalescence position is given by the product of the Boltzmann weights, $\exp(-\beta[\phi_L(x)+\phi_R(x)])$, which is a position-independent constant due to the cancelation of the position dependence of the two opposite linear potentials. This simple picture breaks down close to the boundaries but otherwise is sufficient to grasp the observed behavior.

The characteristic (mean) coalescence times in Fig. 4 cross over from the ff behavior for $f\gg 1$, proportional to the inverse of the force $\tau\simeq 1/f$ [using the natural boundary condition in the calculation of Ref. 31 gives $\tau=(y_0-x_0)/(4f)$, cf. Eq. (9) therein] to the thermal Arrhenius/Kramers-like barrier crossing proportional to the exponential of the barrier height $\tau\simeq\exp(2|f|)$ for $f\ll -1$. Thus, all the results are plausible and can be qualitatively rationalized based on simple physical arguments.

In the rest of this section we focus on a more detailed study of the two limiting cases, the almost ff $f\gg 1$ and the large barrier $f\ll -1$. For these limiting cases we obtain analytical results from relatively simple assumptions which compare quantitatively well with the full solution. We start with the ff case where we assume that the large drift toward coalescence dominates the dynamics so that the reflecting boundary conditions can be safely neglected since the typical realizations/trajectories of the stochastic process never reach them. The validity of this assumption depends on the initial conditions and we expect it to be good enough for the walkers starting from initial positions x_0,y_0 satisfying $\min(x_0, 1-y_0)\gtrsim 1/f$, i.e., far enough from the boundaries. This, indeed, turns out to be the case; see below and compare the corresponding results in Figs. 3(a) and 3(b).

If the reflecting boundary conditions are neglected the two-walker Fokker–Planck equation (30a) can be solved by separation of variables. In particular, the reformulation of Eq. (30a) together with conditions (30f) and (30g) in terms of the center-of-mass $(x+y)/2$ and relative $y-x$ variables leads to a separable problem which can be easily solved since the center-of-mass coordinate (cms) just performs a free diffusion while the relative coordinate (rel) satisfies equations analogous to those in Ref. 31 [Eq. (7) with $c=0$ therein]. Following a derivation similar to that in Sec. V A we arrive at (for details the reader is referred to an upcoming publication[64])

$$\varrho_{\mathrm{ff}}(x;t|x_0,y_0) = P_{\mathrm{cms}}\left(\frac{x+y}{2};t\middle|\frac{x_0+y_0}{2}\right)\pi_{\mathrm{rel}}(t|y_0-x_0), \tag{66}$$

with

$$P_{\mathrm{cms}}(u;t|u_0) = \frac{1}{\sqrt{2\pi t}}\exp\left(-\frac{(u-u_0)^2}{2t}\right) \tag{67}$$

being the free diffusion propagator of the center-of-mass coordinate (see Refs. 65 and 66) and

$$\pi_{\mathrm{rel}}(t|y_0-x_0) = \frac{y_0-x_0}{\sqrt{8\pi t^3}}\exp\left(-\frac{(y_0-x_0-4ft)^2}{8t}\right) \tag{68}$$

the first passage time probability density for the relative coordinate to reach the origin [compare with Eq. 8 in Ref. 31 and the general discussion in Ref. 57]. Using $\rho(x|x_0,y_0)=\int_0^\infty dt\,\varrho(x;t|x_0,y_0)$ and the identity $\int_0^\infty dt\exp(-a^2/(2t)-2b^2t)/t^2=4bK_1(2ab)/a$ for $a,b>0$ [$K_n(x)$ is the modified Bessel function of the second kind of order n] we finally obtain for the coalescence position probability density in the ff limit

$$\rho_{\mathrm{ff}}(x|x_0,y_0) = \frac{f(y_0-x_0)\exp[f(y_0-x_0)]K_1(2fr)}{\pi r}, \tag{69}$$

with

$$r \equiv \sqrt{(x-(x_0+y_0)/2)^2+(y_0-x_0)^2/4}.$$

This function is plotted in Figs. 3(a) and 3(b) for the two different initial conditions for force $f=10$. We can see a rather good agreement between the asymptotic formula (69) and the full result in Fig. 3(b). The situation is much worse in Fig. 3(a) although even there the correspondence is qualitatively quite acceptable. As mentioned above, the reason for the success or failure of the approximation is determined by the initial conditions. Indeed, Fig. 3(b) corresponds to $f\min(x_0,1-y_0)=f(1-y_0)=1$ where the approximation is expected to become valid while in Fig. 3(a) the walkers start out right at the boundaries and only an extremely high value of the force could prohibit the walkers from occasionally bumping into the boundaries, especially at the very beginning. Thus, in the case with $x_0=0$, $y_0=1$ the above approximation is expected to become quantitatively accurate only for very high values of f—numerical estimates reveal that an agreement comparable with that of Fig. 3(b) is not achieved until about $f\gtrsim 40$. This supports a heuristic guess that the accuracy of the asymptotic analytic expression crosses over from $\exp[-f\min(x_0,1-y_0)]$ to $1/f$ when the minimum equals zero.

Now, we turn to the opposite limit of large barrier, i.e., the $f\ll -1$ case. In particular, we want to derive asymptotic expressions for the characteristic coalescence time (which is identified from the results of the full theory as $\tau=1/|\Lambda_0|$, cf. the end of Sec. IV C and below) and the coalescence position probability density. Clearly, this limit is dominated by the lowest eigenvalue and eigenfunction of the two-walker problem which means by the two lowest eigenvalues and eigenfunctions of the auxiliary one-walker problem. Thus we can write, under assumptions of large barrier (lb) and generic initial conditions (to be specified in more detail below), for $P_{\mathrm{lb}}(x,y;t|x_0,y_0)$,

$$\begin{aligned}P_{\mathrm{lb}}(x,y;t|x_0,y_0) \simeq\ & e^{f(x-y)}e^{-f(x_0-y_0)}e^{(\lambda_0+\lambda_1)t}\\ &\times[\psi_0(x)\psi_1(y)-\psi_0(y)\psi_1(x)]\\ &\times[\psi_0(x_0)\psi_1(y_0)-\psi_0(y_0)\psi_1(x_0)],\end{aligned} \tag{70}$$

where $\lambda_{0,1}\simeq\pm 4f^2e^{-|f|}$ are the lowest eigenvalues satisfying Eq. (54a) and $\psi_{0,1}(x)$ are the corresponding eigenfunctions given by (these are exact expressions for all $f<-2$)

$$\begin{aligned}\psi_0(x) &= \sqrt{\frac{2\lambda_0}{\lambda_0+2|f|}}\cosh[\sqrt{f^2+\lambda_0}(x-1/2)],\\ \psi_1(x) &= \sqrt{\frac{2|\lambda_1|}{\lambda_1+2|f|}}\sinh[\sqrt{f^2+\lambda_1}(x-1/2)].\end{aligned} \tag{71}$$

Using these expressions we can study the dependence on the initial conditions and clarify the regime in which the assumption about the dominance of the lowest eigenmode is valid. Utilizing that in the limit $f\ll -1$ the eigenvalues satisfy $|\lambda_1|\simeq\lambda_0\ll|f|$ we obtain

$$\begin{aligned}&e^{-f(x_0-y_0)}[\psi_0(x_0)\psi_1(y_0)-\psi_0(y_0)\psi_1(x_0)]\\ &\quad\simeq\frac{\lambda_0}{|f|}e^{-f(x_0-y_0)}\sinh[|f|(y_0-x_0)]\\ &\quad=\frac{\lambda_0}{2|f|}(1-\exp[-2|f|(y_0-x_0)]).\end{aligned} \tag{72}$$

Since by assumption $y_0>x_0$ we see that the evolution depends only exponentially weakly on the initial conditions so that for $y_0-x_0\gg 1/|f|$ the evolution is essentially independent of the initial conditions as expected in the high-barrier limit. Indeed, the above condition just says that the walkers should start out well separated so that the barrier between them is still large (in dimensionless units). In such a case the dynamics is independent of the detailed initial condition or, more precisely, it depends on it only exponentially weakly which can safely be neglected. This is the regime in which the lowest eigenmode theory of Eq. (70) is sufficient as we will demonstrate below.

If we calculate $\varrho(x;t|x_0,y_0)$ from Eq. (59) in the limit $f\ll -1$, $y_0-x_0\gg 1/|f|$ we obtain

$$\begin{aligned}\varrho_{\mathrm{lb}}(x;t|x_0,y_0) &\simeq \frac{\lambda_0}{|f|}e^{(\lambda_0+\lambda_1)t}[\psi_0(x)\psi_1'(x)-\psi_1(x)\psi_0'(x)]\\ &\simeq \frac{\lambda_0^2}{|f|}e^{(\lambda_0+\lambda_1)t}\left\{1-\frac{\lambda_0}{2f^2}\cosh[2|f|(x-1/2)]\right\}.\end{aligned} \tag{73}$$

Now integrating over time and taking into account that $\lambda_0\simeq 4f^2e^{-|f|}$ and $\lambda_0+\lambda_1\equiv\Lambda_0\simeq -16f^2(|f|-1)e^{-2|f|}<0$ we get for $\rho(x|x_0,y_0)=\int_0^\infty dt\,\varrho(x;t|x_0,y_0)$

$$\rho_{\rm lb}(x|x_0,y_0) \simeq \frac{|f|}{|f|-1}(1-2e^{-|f|}\cosh[2|f|(x-1/2)])$$
$$= \frac{1}{1-1/|f|}(1-e^{-2|f|x}-e^{-2|f|(1-x)}). \quad (74)$$

This clarifies that $\rho_{\rm lb}(x|x_0,y_0)$ is properly normalized to 1 within exponential precision, $\int_0^1 dx\rho_{\rm lb}(x|x_0,y_0)=1+\mathcal{O}(e^{-2|f|}) \simeq 1$, which finally proves the self-consistency of the lowest eigenmode approximation. The curves for the case $f=-10$ in Figs. 3(a) and 3(b) calculated by the full theory are practically indistinguishable from that given by the approximate expression (74) (which is an explicit illustration of the initial-condition independence). In a straightforward manner it also follows that the mean coalescence time is given by the inverse lowest eigenvalue $1/|\Lambda_0|$ since for $\varrho_{\rm lb}(x;t|x_0,y_0)$ in the separable form of Eq. (73) and due to the above normalization condition one immediately gets

$$\tau_{\rm lb}(x_0,y_0) \simeq \int_0^\infty dt\, t\int_0^1 dx\varrho_{\rm lb}(x;t|x_0,y_0)$$
$$= \int_0^\infty dt\, te^{-|\Lambda_0|t}|\Lambda_0|\int_0^1 dx\rho_{\rm lb}(x|x_0,y_0)$$
$$\simeq \frac{1}{|\Lambda_0|} \simeq \frac{e^{2|f|}}{16f^2(|f|-1)}, \quad (75)$$

independent of the initial conditions. All approximate equalities hold up to exponentially small corrections of order $e^{-2|f|}$, which are negligible for $f \ll -1$.

C. Summary

We have in Secs. III–V set up an approximate Fokker–Planck equation scheme for the full problem of two interfaces moving in a block DNA stretch with a barrier region separating two soft zones. While the full problem can be viewed as two discrete random walkers in different potentials with an imposed vicious boundary condition, the approximate Fokker–Planck equation describes two continuous random walkers in opposite linear potentials keeping the imposed viciousness condition. The four assumptions leading to the Fokker–Planck equation were introduced in Sec. III and their validity will be discussed favorably in Sec. VII when we compare the Fokker–Planck results with the direct evaluation of the full discrete problem using the master equation approach presented in Sec. VI.

The main outcome of the Fokker–Planck approach are the general results for the coalescence time density $\pi(t)$ (examples will be shown in Sec. VII) and the numerical and analytic expressions for the mean meeting time τ (shown in Fig. 4) as well as the probability density ρ for the meeting position (shown in Fig. 3). All results are expressed through the dimensionless force f, which in terms of the parameters from an experimental setup reads

$$f = \frac{N(u_b-1)}{u_b+1}, \quad (76)$$

making comparison with values obtained from experiments straightforward.

Finally, a remark on why one should consider the continuous approach over the complete, discrete master equation approach is in order. Namely, for DNA stretches of length N the discrete master equation approach involves diagonalization of matrices of the order $N^2\times N^2$, setting computational limitations on N, and a new diagonalization is needed for each parameter set. The Fokker–Planck approach may therefore provide additional insight for very long DNA stretches as we showed here by discussing the physical quantities of meeting position and meeting time.

VI. COMPLETE DISCRETE APPROACH: THE MASTER EQUATION

In this section, we develop a master equation framework for the bubble coalescence. In contrast to the previous treatment, we explicitly allow the soft zones to zip close from the two ends of the barrier region. It will turn out that in some cases this only has a minor effect. A detailed comparison with the continuum Fokker–Planck equation approximation is shown in Sec. VII.

We consider the same segment of double-stranded DNA with $M=N_L+N+N_R$ internal bps, clamped at both ends according to Fig. 1. However, in contrast to the approximations imposed in the Fokker–Planck approximation, we now allow for explicit closure of the soft regions, necessitating the consideration of a sequence dependence of the local DNA stability, in contrast to the previous discussion, where we assumed that the two bubble domains always remain open.

Note that in this section our notation differs from the scheme introduced above, first, in order to keep the notation of this section consistent with previous references on the same method,[21,24,25,35,39] and second, to be able to incorporate zipping/unzipping of bps in the soft zones, as well. We denote by $x_L=X+N_L$ ($x_R=Y+N_L$) the position of the rightmost (leftmost) open bp in the left (right) open region, see Fig. 1, where $x_{L,R}\in[0,M]$. The positions x_L and x_R of the two zipper forks are stochastic variables and the aim is to understand how these variables evolve in time without taking the continuum limit and using the approximations introduced in Sec. III. We note that an equivalent set of variables are x_L and the clamp size m, that are related through

$$m = x_R - x_L. \quad (77)$$

In the master equation formulation below we will use x_L and m as the dynamic variables, and for completeness we state the transition rates, Eqs. (4), (5), (7), (8), (12), and (13) expressed in the new variables: The reflecting boundary conditions are

$$t_L^-(x_L=0,m) = t_R^+(x_L,m=M-x_L) = 0. \quad (78)$$

Once the clamp is completely unzipped, i.e., the state $m=0$ is reached, we assume that the clamp will not be able to reform for a long time, and we impose the absorbing conditions

TABLE I. Properties of the transfer coefficients. The quantity Δx_L (Δx_R) denotes the change in fork position x_L (x_R) under the action of the transfer coefficients. Similarly Δm denotes the change in clamp size m.

	Δx_L	Δx_R	Δm	Eq.
$t_L^+(x_L,m)$	1	0	−1	(80)
$t_R^-(x_L,m)$	0	−1	−1	(81)
$t_L^-(x_L,m)$	−1	0	1	(82)
$t_R^+(x_L,m)$	0	1	1	(83)

$$t_L^-(x_L,m=0)=t_R^+(x_L,m=0)=0. \tag{79}$$

The transition rates at the interior of the DNA stretch are

$$t_L^-(x_L,m)=\tfrac{1}{2}\mathcal{K}(x_L), \tag{80}$$

$$t_R^+(x_L,m)=\tfrac{1}{2}\mathcal{K}(M-m-x_L), \tag{81}$$

$$t_L^+(x_L,m)=\tfrac{1}{2}\mathcal{K}(x_L+1)u(x_L+1)s(m), \tag{82}$$

$$t_R^-(x_L,m)=\tfrac{1}{2}\mathcal{K}(M-m-x_L+1)u(x_R-1)s(M-m-x_L), \tag{83}$$

where $x_R=x_L+m+1$, $s(q)=\{(q+1)/(q+2)\}^c$, and $\mathcal{K}(q)=kq^{-\mu}$. The properties of the four transfer coefficients above are summarized in Table I.

Denote by $P(x_L,m,t|x_L',m')$ the conditional probability to find the system in state (x_L,m) at time t given the initial condition (x_L',m') at time $t=0$. With the shorthand notation $P(x_L,m,t)=P(x_L,m,t|x_L',m')$ the dynamics is described by the master equation

$$\begin{aligned}\frac{\partial}{\partial t}P(x_L,m,t)&=t_L^+(x_L-1,m+1)P(x_L-1,m+1,t)\\&+t_L^-(x_L+1,m-1)P(x_L+1,m-1,t)\\&-[t_L^+(x_L,m)+t_L^-(x_L,m)]P(x_L,m,t)\\&+t_R^+(x_L,m-1)P(x_L,m-1,t)\\&+t_R^-(x_L,m+1)P(x_L,m+1,t)\\&-[t_R^+(x_L,m)+t_R^-(x_L,m)]P(x_L,m,t).\end{aligned} \tag{84}$$

This equation states that the probability for the clamp size can change in eight different ways: the terms with a plus sign correspond to jumps *to* the state $\{x_L,m\}$, and the terms with a minus sign correspond to jumps *from* the state $\{x_L,m\}$.

A standard approach to the master equation (84) is the spectral decomposition[56,57]

$$P(x_L,m,t)=\sum_p c_p(x_L',m')Q_p(x_L,m)\exp(-\eta_p t) \tag{85}$$

in terms of the eigenvalues η_p and eigenvectors $Q_p(x_L,m)$. The expansion coefficients $c_p(x_L',m')$ are obtained from the initial condition. As in Sec. V, we will assume the system initially to be in the state where all bps in the soft zone are broken and all bps in the barrier region are closed.

The eigenvalue equation corresponding to Eq. (84) becomes

$$\begin{aligned}&t_L^+(x_L-1,m+1)Q_p(x_L-1,m,t)+t_L^-(x_L+1,m-1)Q_p(x_L+1,m-1,t)\\&\quad-[t_L^+(x_L,m)+t_L^-(x_L,m)]Q_p(x_L,m,t)+t_R^+(x_L,m-1)Q_p(x_L,m-1,t)\\&\quad+t_R^-(x_L,m+1)Q_p(x_L,m+1,t)-[t_R^+(x_L,m)+t_R^-(x_L,m)]Q_p(x_L,m,t)\\&=-\eta_p Q_p(x_L,m,t).\end{aligned} \tag{86}$$

The eigenvectors satisfy the orthogonality relation[57]

$$\sum_{m=1}^{M}\sum_{x_L=0}^{M-m}\frac{Q_p(x_L,m)Q_{p'}(x_L,m)}{P_r^{\mathrm{eq}}(x_L,m)}=\delta_{pp'}, \tag{87}$$

where

$$P_r^{\mathrm{eq}}(x_L,m)=\mathcal{Z}(x_L,m)/\mathcal{Z}, \tag{88}$$

with

$$\mathcal{Z}=\sum_{x_L,m}Z(x_L,m) \tag{89}$$

being the partition coefficient in the variable x_L and m (neglecting a common σ_0^2 factor, where σ_0 is the bubble initiation parameter) and where

$$\begin{aligned}\mathcal{Z}(x_L,m)=&\prod_{\tilde{x}_L=0}^{x_L}u(\tilde{x}_L)\prod_{\tilde{x}_R=m+x_L+1}^{M+1}u(\tilde{x}_R)(1+x_L)^{-c}\\&\times(M-m-x_L+1)^{-c},\end{aligned} \tag{90}$$

see Eqs. (1)–(3). From the eigenvalues and eigenvectors of Eq. (86) any quantity of interest may be constructed. In Sec. VI A we calculate the coalescence time density. How to set up the master equation for numerical purposes is presented in detail in Appendix B. Alternatively, the master equation (84) can be solved by direct stochastic simulations such as the Gillespie algorithm introduced in Appendix C and used in Sec. VII.

A. Coalescence time density

The (survival) probability that the absorbing boundary at $m=0$ has not yet been reached up to time t is

$$\mathcal{S}(x_L',m',t) = \sum_{m=1}^{M} \sum_{x_L=0}^{M-m} P(x_L,m,t|x_L',m'). \tag{91}$$

The probability that the absorbing boundary is reached within the time interval $[t,t+dt]$ (namely, the coalescence time density corresponding to the first passage problem) is

$$\rho(x_L',m',t)dt = \mathcal{S}(x_L',m',t) - \mathcal{S}(x_L',m',t+dt)$$
$$= -\left(\frac{\partial}{\partial t}\mathcal{S}(x_L',m',t)\right)dt. \tag{92}$$

This expression is positive, as $\mathcal{S}$ decreases with time. To express the coalescence time density in terms of the eigenvalues and eigenfunctions, we introduce the eigenmode expansion (85) into Eq. (92), yielding

$$\rho(x_L',m',t) = \sum_p \eta_p c_p(x_L',m')\exp(-\eta_p t), \tag{93}$$

with coefficients

$$c_p(x_L',m') = \frac{Q_p(x_L',m')}{P_r^{\mathrm{eq}}(x_L',m')} \sum_{m=1}^{M} \sum_{x_L=0}^{M-m} Q_p(x_L,m). \tag{94}$$

We have above made use of the orthonormality relation (87) in order to express $c_p(x_L',m')$ in terms of the initial probability density $P(x_L,m,0|x_L',m')$ and used the fact that this general initial condition takes the explicit form $P(x_L,m,0|x_L',m') = \delta_{x_L,x_L'}\delta_{m,m'}$. Equation (93) is the discrete counterpart of the continuous result derived in Ref. 58 and expresses the coalescence time density (for any given initial condition, specified by m' and x_L') in terms of the eigenvalues and eigenvectors of Eq. (86).

VII. COMPARISON BETWEEN THE FULL MASTER EQUATION AND THE FOKKER–PLANCK APPROXIMATION

In this section, we investigate the validity of the assumptions presented in Sec. III leading to the Fokker–Planck continuum approximation. This is done by comparing the results for the coalescence time densities, $\pi(t)$, with the results obtained from the full discrete master equation approach. In the next section, Sec. VIII, we discuss the relevance for biological experiments.

In all examples below the two walkers move between $[0,N]$ in the Fokker–Planck description and between $[-N_L,N+N_R]$ in the master equation setup. That is, in the master equation approach we explicitly allow zipping of bps in the two soft zones. As initial conditions we use $X_0=0$ and $Y_0=N$ throughout this section.

A. The continuum approximation

The continuum assumption (iv) implies that the inherently discrete nature of the DNA structure—both in terms of stacking and hydrogen bonds—can be approximated by the diffusive behavior of two continuous variables. To get the Fokker–Planck description one has to consider the limit $a \to 0$ with a being the length between effective bonds in the bps. In practice this limit is obtained by

$$\frac{\text{bond distance}}{\text{total segment length}} \to 0,$$

i.e., by increasing the width of the barrier region. The question is addressed in a setup with open soft zones, i.e., $u_s \gg 1$, and varying barrier lengths, N. To have perfectly reflecting boundary conditions we set $N_L=N_R=0$ in the master equation setup.

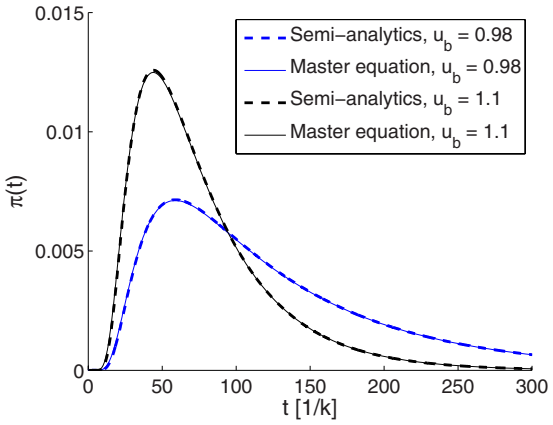

FIG. 5. (Color online) Coalescence time probability density for barrier width $N=20$ and temperature below and above T_b, with $u_b=0.98$ and $u_b=1.1$, respectively. To exclude other effects, the lengths of the soft zones are zero, $N_L=N_R=0$, i.e., these are assumed to be always open, and $c=\mu=0$.

Figure 5 shows that it requires a relatively small number of bps ($\sim$20) before the continuum approximation is reasonable independently of the temperature. Barrier regions of this length are in principle accessible experimentally so the continuum approximation appears to be well justified.

B. Open soft zones

The first assumption, (i), states that the soft zones are always open, i.e., the random walking zipper forks are reflected at the interfaces between the barrier region and the soft zones. To eliminate other effects than the effect of the introduction of the reflecting boundary conditions at the ends of the barrier region, we consider a DNA stretch of length 25, so that the continuum approximation is justified, and set $c=\mu=0$ in order to exclude effects originating from the entropy factor and the hook exponent. We compare this to the results from the full master equation including the soft zones.

Figure 6 shows the coalescence time density $\pi(t)$ for varying lengths of the soft zones for temperatures above and below the melting temperature of the barrier region. Apparently, the length dependence is rather weak as long as the soft zones serve as sufficiently hard boundaries, i.e., for large enough $u_s \gtrsim 5$. For smaller u_s the soft-zone-length dependence is relevant as the two bubble corners venture more frequently into the soft zones. This, however, implies the breaking of the assumptions for the applicability of our Fokker–Planck description as revealed in the figure. Thus, the length of the soft zones itself is not important if u_s is large enough and $c=\mu=0$. A systematic assessment of these conditions is given below.

Figure 7 shows $\pi(t)$ as a function of u_s for $N_L=N_R=20$, i.e., the soft zones are so long that the two forks essentially

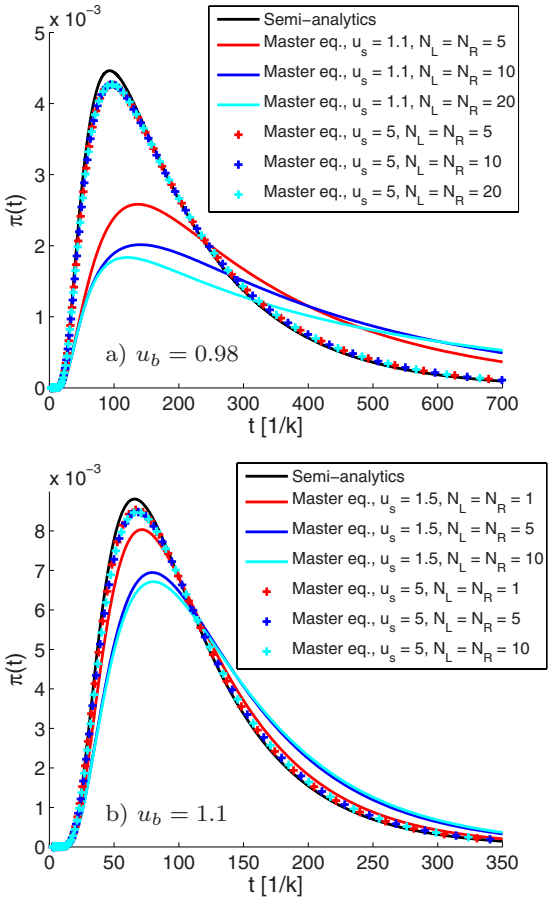

FIG. 6. (Color online) Coalescence time density for varying soft-zone lengths and fixed length of the barrier $N=25$. Included are plots for $u_b=0.98$ (upper) and $u_b=1.1$ (lower), i.e., for temperatures below and above the melting temperature of the barrier, respectively. Furthermore, $c=\mu=0$.

never reach the outer clamps. That this is indeed the case can be qualitatively investigated using the Gillespie scheme presented in Appendix C, giving access to real-time trajectories of the two random walkers in a potential landscape including both the barrier region and the soft zones. This is shown in Fig. 8 which confirms that excursions into the soft zones are progressively suppressed with increasing u_s. In Fig. 7 we have only included the case $u_b<1$, i.e., when the barrier

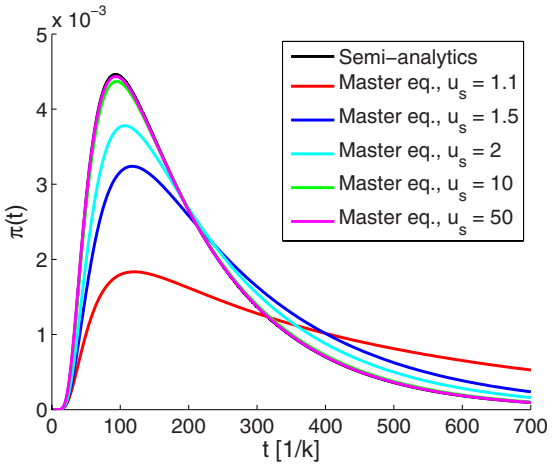

FIG. 7. (Color online) Coalescence time density for $N=25$, $N_L=N_R=20$, $u_b=0.98$, and different values of u_s. Furthermore, $c=\mu=0$.

region indeed acts as a barrier, and consequently the effect of the soft zones is more pronounced. The discrepancies between the Fokker–Planck and the master equation approach become distinct for $u_s \simeq 1$ whereas for $u_s \gtrsim 5$ the agreement between the approaches becomes reasonable. A difference between u_s and u_b of this magnitude can indeed be achieved in realistic experimental setups, as shown in Sec. VIII.

C. Loop entropy and hook factors

The general rates defined in Sec. II include both the entropy loss factor and the hook factor. Both depend on the length q of the bubble, and for both their relative influence diminishes for increasing bubble lengths. In the Fokker–Planck description both factors are omitted, which are assumptions (ii) and (iii). These assumptions are valid for long bubbles, which can be obtained by having long soft zones and keeping the temperature far above the melting temperature of the soft zones, i.e., $u_s \gg 1$.

Figure 9 studies the effect of the loop exponent c on the coalescence time density $\pi(t)$ for a fixed length of the barrier region, $N=15$, and varying the lengths of the soft zones. Even for soft zones of length $N_L=N_R=30$ there is a substantial influence of the loop entropy factor, making long soft zones a requirement in experimental realizations which should agree reasonably with the Fokker–Planck approach.

The hook factor leads to a decrease in the rate constant k, $k \to kq^{-\mu}$ where q is the length of a given bubble, so introducing the hook exponent $\mu>0$ leads to a considerable and bubble-length-dependent decrease in the transition rates, see Fig. 10. However, for sufficiently large bubbles, the rate is roughly constant and introducing a renormalized rate constant $\tilde{k}=kL^{\mu}$ can compensate for this effect. Here L is a characteristic bubble size. If the temperature is kept well above the melting temperature of the soft zones, and the length of the soft zones is much longer than the barrier, L is well approximated by the length of the soft zones, $L \simeq N_{L/R}$. In Fig. 10 we illustrate the effect of a renormalized rate constant (with $L=53$ being the best fit) together with the standard rate coefficient $k=1$. In conclusion, both the influence of the loop entropy factor and the hook exponent can be eliminated by keeping the length of the soft zones sufficiently long and using a renormalized value for the rate constant k.

VIII. RELEVANCE FOR SINGLE MOLECULE EXPERIMENTS

Relevant for the separation of statistical weights are the empirical relations[8]

$$T_m^{\mathrm{AT}} = (355.55 + 7.95 \ln[\mathrm{Na}^+])K, \tag{95a}$$

$$T_m^{\mathrm{GC}} = (391.55 + 4.98 \ln[\mathrm{Na}^+])K, \tag{95b}$$

which give the melting temperatures of GC and AT pairs in terms of the (intermediate) salt concentration in the solvent obtained from melting experiments.[67] Note that the value of T_m^{AT} stems from an average over all possible combinations of AT and TA bps;[8] if only TA/AT and AT/TA pairs are interchangeably used the value of T_m^{AT} can be lowered further.

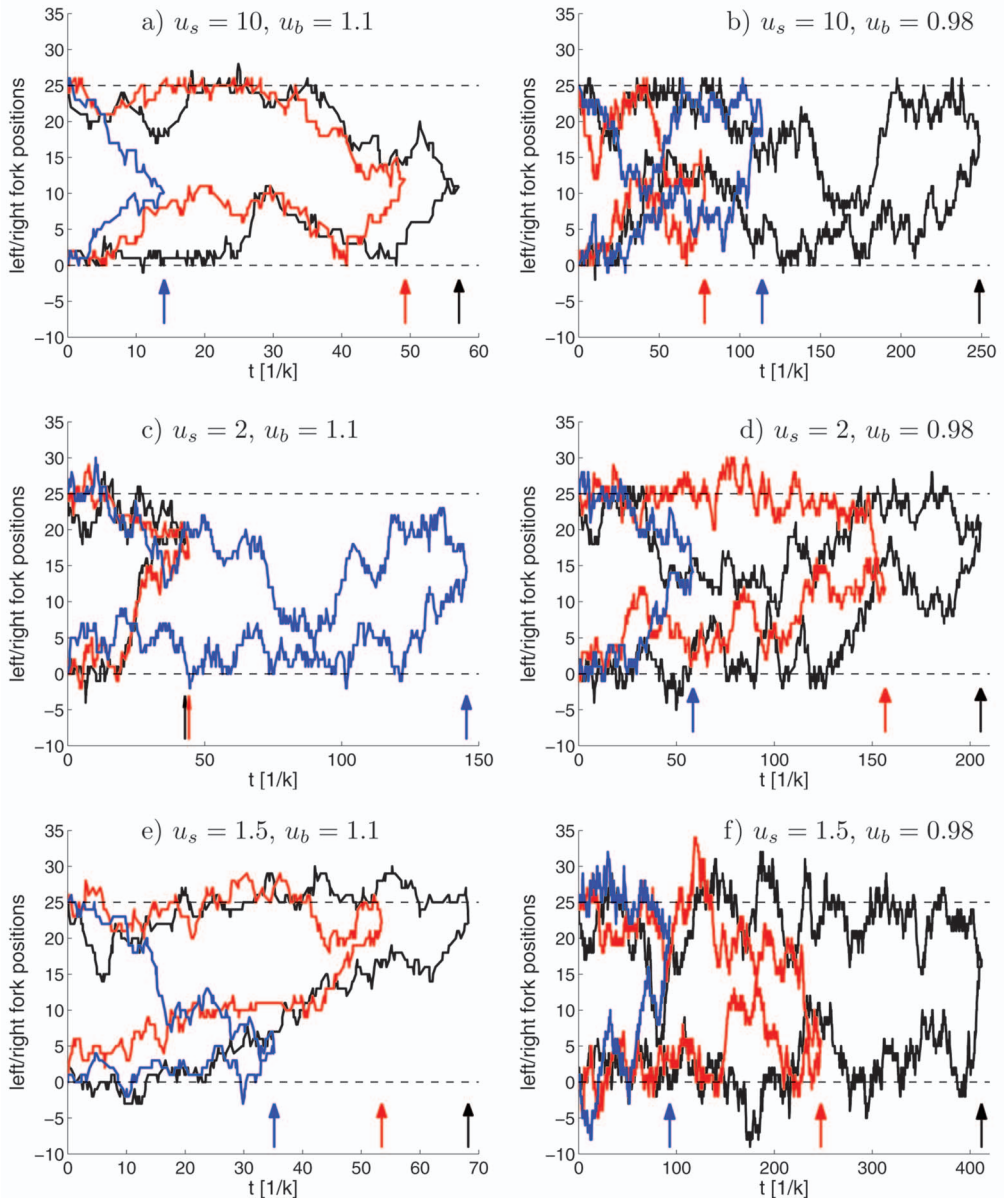

FIG. 8. (Color) Single trajectories for the two zipper forks, based on the Gillespie (Monte Carlo) algorithm presented in Appendix C. The values of u_b and u_s are stated in the figures. The other parameters are $N_L=N_R=20$, $N=25$, and $c=\mu=0$. The dashed horizontal lines mark the boundaries between the barrier region and the soft zones, and the arrows indicate the ends of the sampled trajectories.

The above relations can be translated into free energy differences by $\Delta G=\Delta S(T_m-T)$, where $\Delta S=-24.85$ cal/mol K.[68] Equations (95a) and (95b) contain contributions from both base stacking and hydrogen bonding and is thus the melting temperature suitable for our situation. It has been shown that the dependence on salt concentration lies in the stacking term[68] and not as previously thought in the hydrogen bonding term.[69] The stacking is a combination of hydrophobic, electrostatic (screening of the negatively charged phosphate groups), and dispersive interactions but there is no apparent consensus on which term is the dominant one.[68] At high salt concentrations $\sim 1M-5M$ the temperature dependence levels off due to a decrease in the hydrophobic effect; with most water molecules tied up in the solvation of ions, the entropy decrease involved in base stacking is small.[70]

The melting temperature of AT bonds has a stronger dependence on salt concentration, so we can increase the ratio $u_{\mathrm{AT}}/u_{\mathrm{GC}}$ by decreasing the salt concentration. A further benefit is a lowering of the melting temperatures, thus enabling experiments well below $T_m^{\mathrm{GC}}\sim 100$ °C, which is the melting temperature when $[\mathrm{Na}^+]=0.1M$, the standard concentration in electrophoresis experiments. High temperatures have practical disadvantages such as formation of air bubbles and increased evaporation of solvent molecules.[70]

Most relevant experiments on DNA have been conducted at $\sim 0.1M$ salt concentration. Those specifically looking at the salt dependence of the melting temperature work in the range $\sim 0.01M-1M$. A conservative estimate of $[\mathrm{Na}^+]=0.01M$ gives $u_{\mathrm{GC}}\sim 1$ and $u_{\mathrm{AT}}\simeq 6$ at $T\simeq 95$ °C, which is sufficient for the Fokker–Planck approximation to be valid.

Concerning the possible length of a DNA construct it should be reasonable to work with segments up to 100–200 bps. In a setup combining fluorescence correlation spectroscopy and fluorescence quenching, as introduced in Ref. 16,

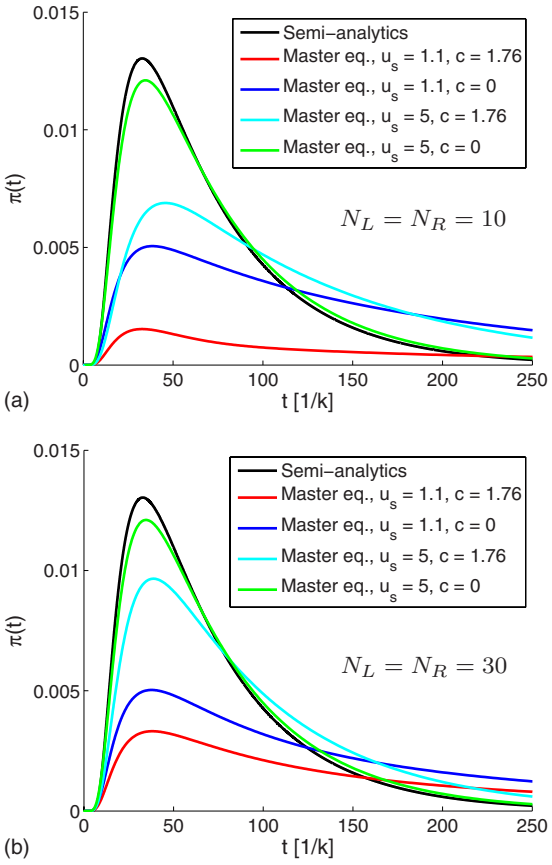

FIG. 9. (Color online) Coalescence time density $\pi(t)$ for fixed barrier width, $N=15$, and varying the length of the soft zones, $N_L=N_R=10$ in (a) and $N_L=N_R=30$ in (b). The other parameters are $u_b=0.98$ and $\mu=0$.

the DNA is free to diffuse around in the solution. In this case the limiting factor is the time it takes for the quencher to diffuse in and out of the confocal volume.

Practically the experimental method of choice may be a dual optical tweezer setup in which the DNA construct, via some handles of double-stranded DNA, is connected to two beads held in place by the tweezers. While this allows to keep the DNA construct in place the force exerted on the chain is relatively small. However, this would allow direct observation of the construct avoiding diffusional correction. The center of the barrier region could be decorated with either a fluorophore-quencher pair or markers such as quantum dots or small gold beads that can be visualized by a microscope. The influence of the attached markers should decrease with longer barrier length. Having this setup in a flow cell, the system could be triggered by flushing in a solution with either different temperature or salt concentration. This can be done relatively quickly.[71] Once the two initial bubbles are thereby created it should be possible to measure the coalescence time calculated herein. Repeating the experiment would produce the distribution of coalescence times, from which important system parameters can be inferred. Decorating the barrier region with several, sufficiently small, markers would, in principle, allow one to measure the position of the coalescence.

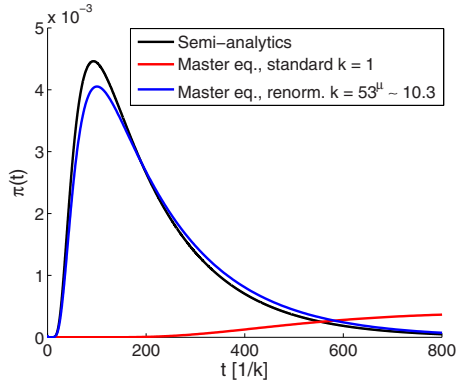

FIG. 10. (Color online) Coalescence time density $\pi(t)$ for nonvanishing hook exponent $\mu=0.588$, $N_L=N_R=50$, and $N=25$ for the barrier case and $u_b=0.98$. The results are shown with and without the renormalized rate constant k, where we have used $\tilde{k}=53^{0.588}\simeq 10.3$. The other parameters are $u_s=10$ and $c=0$.

IX. FINAL CONCLUSIONS

Single molecule techniques give us increasing insight into the behavior, equilibrium, and dynamics of biopolymers. Of particular and outstanding interest is DNA due to its importance in biological contexts as well as its role as a model biopolymer. In order to extend our knowledge about the biological function of DNA it is crucial to quantify and understand the denaturation behavior of DNA at the single molecule level, its sequence dependence, and, ultimately, its relevance to genetic processes such as transcription initiation. A major question hereby concerns the dynamics of transient DNA denaturation bubbles.

While first single molecule fluorescence correlation experiments have demonstrated the feasibility of monitoring the fluctuations of a single bubble, some questions remain about the model system and the explicit setup used in these experiments. In particular, the obtained time scales for bp zipping and unzipping as well as the influence of the attached fluorophore-quencher pair remain under debate.

Here we suggest an alternative model for accessing DNA stability parameters and bp (unzipping) constants: in our model system DNA bubbles in two AT-rich regions are formed and separated by a more stable GC-rich barrier region. The coalescence behavior of the two DNA bubbles across the barrier region is then studied. We show that the stability parameters for bubble and barrier regions can indeed be chosen sufficiently different to allow preparation of the DNA construct in the proposed fashion by the proper adjustment of temperature and/or salt concentration. Once coalesced the newly created single bubble is stabilized against immediate reclosure of the barrier region both dynamically and due to the release of the boundary free energy corresponding to one cooperativity factor σ_0. Appropriate fluorophore-quencher tagging of the barrier region bps should therefore allow for the direct observation of the bubble merging dynamics.

Apart from the relevance of the investigated system for understanding the dynamics of DNA and its biological function, the mathematical description presented here is of interest for its own sake as it corresponds to a previously unsolved case of two vicious random walkers in opposite linear

potentials. We established the solution of this problem by solving a bivariate Fokker–Planck equation (continuum limit of the discrete master equation description) analytically. In a careful analysis we showed under what conditions the Fokker–Planck approach is valid and what deviations one would expect for realistic systems. Furthermore, the analytic results were explained using qualitative arguments and corroborated using stochastic simulations.

ACKNOWLEDGMENTS

The work of T.N. is a part of the research plan MSM 0021620834 financed by the Ministry of Education of the Czech Republic and was also partly supported by Grant No. 202/08/0361 of the Czech Science Foundation. T.A. acknowledges funding from the Knut and Alice Wallenberg foundation. R.M. acknowledges the Natural Sciences and Engineering Research Council (NSERC) of Canada and the Canada Research Chairs program for support. This work was started at CPiP 2005 (Computational Problems in Physics, Helsinki, May 2005) supported by NordForsk, Nordita, and Finnish NGSMP. We gratefully acknowledge very helpful discussions with Oleg Krichevsky.

APPENDIX A: CALCULATION OF $\tilde{p}(x;t|x_0)$ VIA LAPLACE TRANSFORM

In this appendix we present a detailed calculation of the single-walker auxiliary density $\tilde{p}(x;t|x_0)$ satisfying Eq. (49) together with the boundary conditions Eqs. (37a) and (37b), as well as the initial condition $\tilde{p}(x;t=0|x_0)=\delta(x-x_0)$. $\tilde{p}(x;t|x_0)$ solves the Schrödinger equation

$$\frac{\partial}{\partial t}\tilde{p}(x;t|x_0)=\left[\frac{\partial^2}{\partial x^2}-f^2\right]\tilde{p}(x;t|x_0), \tag{A1}$$

that, after a Laplace transform and some rearrangement, becomes

$$\left[\frac{\partial^2}{\partial x^2}-k(z)^2\right]\tilde{p}(x;z|x_0)=-\delta(x-x_0), \tag{A2}$$

with $k(z)=\sqrt{z+f^2}$ (we skip the explicit z dependence in the formulas from now on).

Consider first the solution of the equation

$$\left[\frac{\partial^2}{\partial x^2}-k^2\right]g(x,x_0)=0 \tag{A3}$$

for $x<x_0$ with boundary condition (37a) and for $x>x_0$ with boundary condition (37b). The solutions are

$$g(x,x_0)=\begin{cases} C_<(\kappa e^{kx}+e^{-kx}), & x<x_0, \\ C_>(e^{kx}+\kappa e^{2k}e^{-kx}), & x>x_0, \end{cases} \tag{A4}$$

with $\kappa\equiv(k+f)/(k-f)$.

The solution of Eq. (A2) can now be found by the ansatz

$$\tilde{p}(x;z|x_0)=\begin{cases} C(\kappa e^{kx}+e^{-kx})(e^{kx_0}+\kappa e^{2k}e^{-kx_0}), & x<x_0, \\ C(\kappa e^{kx_0}+e^{-kx_0})(e^{kx}+\kappa e^{2k}e^{-kx}), & x>x_0, \end{cases} \tag{A5}$$

such that

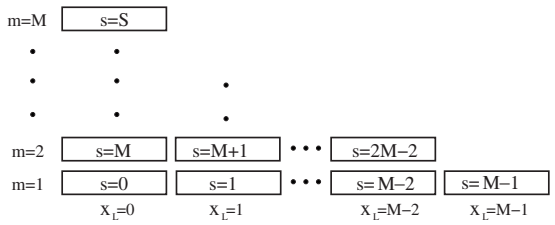

FIG. 11. Enumeration scheme for the numerical analysis: The two-dimensional grid points (x_L,m) are replaced by a one-dimensional running variable s. See text for details.

$$\tilde{p}(x;z|x_0)=C\{\kappa(e^{k(x+x_0)}+\kappa e^{2k}e^{-k|x-x_0|}) + e^{k|x-x_0|}+\kappa e^{2k}e^{-k(x+x_0)}\}, \tag{A6}$$

where C is determined from the jump condition by integrating Eq. (A2),

$$\begin{aligned} -1 &= \lim_{\epsilon\to 0^+}\int_{x_0-\epsilon}^{x_0+\epsilon}\left[\frac{\partial^2}{\partial x^2}-f^2\right]\tilde{p}(x;z|x_0)dx \\ &= \lim_{\epsilon\to 0^+}\left\{\frac{\partial}{\partial x}\tilde{p}(x;z|x_0)|_{x_0+\epsilon}-\frac{\partial}{\partial x}\tilde{p}(x;z|x_0)|_{x_0-\epsilon}\right\} \\ &= 2kC(1-\kappa^2e^{2k}). \end{aligned} \tag{A7}$$

Here, it has been used that $\tilde{p}(x;z|x_0)$ is continuous.

APPENDIX B: IMPLEMENTATION OF THE DISCRETE MASTER EQUATION

To solve the eigenvalue equation (86) by a numerical scheme, it is convenient to replace the two-dimensional grid points (x_L,m) by a one-dimensional coordinate s counting all lattice points, compare with Ref. 35. We choose the enumeration illustrated in Fig. 11. From this figure we notice that $m\in[1,M]$ and $x_L\in[0,M-m]$. An arbitrary s point can be obtained from a specific (x_L,m) according to

$$s=(m-1)M-\frac{(m-1)(m-2)}{2}+x_L. \tag{B1}$$

From this relation we notice that the maximum s value is

$$S=\max\{s\}=\frac{M(M+1)}{2}-1, \tag{B2}$$

i.e., the size of the relevant W matrix (see below) scales as $M^2/2$. Expression (B1) allows us to change the transfer coefficients to the s variable, $t^{\pm}_{L/R}(x_L,m)\to t^{\pm}_{L/R}(s)$, using the explicit expressions (80)–(83) for the transfer coefficients, together with the boundary conditions in Eqs. (78) and (79). From Eq. (B1) and Fig. 11 we notice that [compare with Eq. (86)]

$$s|^{m+1}_{x_L-1}=s|^{m}_{x_L}+M-m \quad \text{for } x_L\geq 1,$$

$$s|^{m-1}_{x_L+1}=s|^{m}_{x_L}-(M-m+1) \quad \text{for } m\geq 2,$$

$$s|^{m-1}_{x_L}=s|^{m}_{x_L}-(M-m+2) \quad \text{for } m\geq 2,$$

$$s|_{x_L}^{m+1} = s|_{x_L}^{m} + M - m + 1$$
$$\text{for } x_L \leq M-(m+1) \quad \text{and} \quad m \leq M-1. \tag{B3}$$

Equation (86) can then be written in matrix form as

$$\sum_{s'} W(s,s')Q_p(s') = -\eta_p Q_p(s), \tag{B4}$$

where explicitly the matrix elements are

$$W(s,s+M-m) = t_L^+(s+M-m) \quad \text{for } s\Cap x_L \geq 1,$$

$$W(s,s-[M-m+1]) = t_L^-(s-[M-m+1])$$
$$\text{for } s\Cap m \geq 2,$$

$$W(s,s-[M-m+2]) = t_R^+(s-[M-m+2])$$
$$\text{for } s\Cap m \geq 2,$$

$$W(s,s+M-m+1) = t_R^-(s+M-m+1)$$
$$\text{for } s\Cap x_L \leq M-(m+1)$$
$$\text{and } m \leq M-1,$$

$$W(s,s) = -(t_L^+(s)+t_L^-(s)+t_R^+(s)+t_R^-(s)), \tag{B5}$$

and the remaining matrix elements are equal to zero. We have introduced the notation $s\Cap$ with the meaning "s is to be taken for." The problem at hand is that of determining the eigenvalues and eigenvectors of the $(S+1)\times(S+1)$ matrix W above. The coalescence time density is then calculated from Eqs. (93) and (94). In terms of the running variable s, see Eq. (B1), and the W matrix defined in Eq. (B5) the detailed balance conditions (10) and (11) become

$$W(s,s')\mathcal{Z}(s') = W(s',s)\mathcal{Z}(s). \tag{B6}$$

The orthogonality relation, Eq. (87), becomes

$$\sum_s \frac{Q_p(s)Q_{p'}(s)}{P_r^{\text{eq}}(s)} = \delta_{p,p'}. \tag{B7}$$

Convenient checks of the numerical results then include the following: (i) The eigenvalues should be real and negative (so that $\eta_p > 0$); (ii) the eigenvectors should satisfy the orthonormality relation, Eq. (B7).

APPENDIX C: STOCHASTIC SIMULATION OF BUBBLE COALESCENCE

In this section we give a brief introduction to the stochastic simulation of DNA breathing; for details we refer to Ref. 36. We apply the Gillespie algorithm introduced in 1976 as a stochastic approach to the study of chemical reactions.[72]

Following the schematic of Fig. 1, we simulate the dynamics of the two zipping forks separating the two initial bubble domains from the barrier region. As each fork can either zip or unzip, the system is described by the four different rates, t_ν^μ, where $\mu \in \{+,-\}$ and $\nu \in \{L,R\}$. Given these rates, we assume that the statistical weight for a given event, $\{\mu,\nu\}$, to occur in a time interval $[t,t+\delta t]$ is $t_\nu^\mu \delta t$. Then the idea of the Gillespie scheme is the following:[72] The probability that nothing happens in the time interval $[t,t+\tau]$ *and* that in the following interval $[t+\tau,t+\tau+d\tau]$ an event of type $\{\mu,\nu\}$ occurs is the so-called reaction probability density

$$P(\tau,\mu,\nu)d\tau = P_0(\tau)t_\nu^\mu d\tau. \tag{C1}$$

To determine the probability $P_0(\tau)$ that no event happens within $[t,t+\tau]$, this interval is divided into K spans of duration $\epsilon = \tau/K$. The probability that no event occurs in the first subinterval $[t,t+\epsilon]$ is then

$$\prod_{\mu,\nu}[1-t_\nu^\mu \epsilon] = 1 - \sum_{\mu,\nu} t_\nu^\mu \epsilon + \mathcal{O}(\epsilon^2). \tag{C2}$$

Treating the remaining intervals similarly produces an expression for P_0,

$$P_0(\tau) = \left[1 - \sum_{\mu,\nu} t_\nu^\mu \epsilon + \mathcal{O}(\epsilon^2)\right]^K$$
$$= \left[1 - \sum_{\mu,\nu} t_\nu^\mu \tau/K + \mathcal{O}(K^{-2})\right]^K. \tag{C3}$$

Taking the limit $K\to\infty$ and reinserting in Eq. (C1), we find the Poissonian law

$$P(\tau,\mu,\nu) = t_\nu^\mu \exp\left(-\sum_{\mu,\nu} t_\nu^\mu \tau\right). \tag{C4}$$

At some given instant of time t, the system is in a certain configuration. The update is performed as follows:

(i) The rates t_ν^μ are calculated according to the configuration.
(ii) A set of random numbers (τ,μ,ν), distributed according to $P(\tau,\mu,\nu)$ in Eq. (C4), is drawn from a generator.
(iii) The time is advanced according to $t\to t+\tau$, and the configuration is updated according to the randomly chosen event μ,ν.

Steps (i)–(iii) are repeated until a specified stop criterion is fulfilled, in our case the merging of the two initial bubbles. We record the stop time and the final configuration, and a new run is initiated using the same initial condition.

Following Ref. 72 we briefly present how random numbers τ and μ can be constructed using numbers drawn from a uniform distribution: Let $P_c(\tau')$ be some continuous probability density function, e.g., $P_c(\tau')d\tau$ is the probability for finding a τ within the interval $[\tau',\tau'+d\tau]$. The associated probability distribution function is then defined as

$$F_c(\tau_0) = \int_{-\infty}^{\tau_0} P_c(\tau')d\tau', \tag{C5}$$

which is the probability of some τ being less than τ_0. To get a random τ according to P_c, given some random number $R \in [0,1]$ drawn from the uniform distribution, we have to invert $F_c(\tau)=R$. Using $P_c(\tau)=\Sigma_{\mu,\nu}P(\tau,\mu,\nu)$ from Eq. (C4), with $\tau>0$, and inverting the expression, we obtain

$$\tau = \frac{1}{\sum_{\mu,\nu} t_\nu^\mu} \ln\left(\frac{1}{R}\right). \tag{C6}$$

Similarly, we determine the appropriate random number for the direction of the "reaction" (zipping/unzipping of left/right zipper fork) following

$$F_d(\mu_0) = \sum_{\nu=1}^{\mu_0} P_d(\nu), \tag{C7}$$

which is the probability of having $\mu \leq \mu_0$. Inversion given some random number $R \in [0,1]$ drawn from the uniform distribution now requires that $F_d(\mu-1) < R < F_d(\mu)$. Using $P_d(\mu) = \int P(\tau,\mu) d\tau$ the random event μ is determined by

$$\sum_{\nu=1}^{\mu-1} r_\nu < R \sum_{\nu=1}^{N} r_\nu \leq \sum_{\nu=1}^{\mu} r_\nu. \tag{C8}$$

[1] J. D. Watson and F. H. C. Crick, Nature (London) **171**, 737 (1953).
[2] C. R. Cantor and P. R. Schimmel, *Biophysical Chemistry* (Freeman, New York, 1980).
[3] A. Kornberg, *DNA Synthesis* (Freeman, San Francisco, 1974).
[4] A. Kornberg and T. A. Baker, *DNA Replication* (Freeman, New York, 1992).
[5] M. D. Frank-Kamenetskii, Phys. Rep. **288**, 13 (1997).
[6] S. G. Delcourt and R. D. Blake, J. Biol. Chem. **266**, 15160 (1991).
[7] R. D. Blake, J. W. Bizzaro, J. D. Blake, G. R. Day, S. G. Delcourt, J. Knowles, K. A. Marx, and J. SantaLucia, Jr., Bioinformatics **15**, 370 (1999).
[8] A. Krueger, E. Protozanova, and M. D. Frank-Kamenetskii, Biophys. J. **90**, 3091 (2006).
[9] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic, New York, 1970).
[10] M. Peyrard, Nat. Phys. **2**, 13 (2006).
[11] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[12] C. Richard and A. J. Guttmann, J. Stat. Phys. **115**, 925 (2004).
[13] E. Yeramian, Gene **255**, 139 (2000); **255**, 151 (2000).
[14] E. Carlon, M. L. Malki, and R. Blossey, Phys. Rev. Lett. **94**, 178101 (2005).
[15] M. Guéron, M. Kochoyan, and J.-L. Leroy, Nature (London) **328**, 89 (1987).
[16] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).
[17] R. Metzler, T. Ambjörnsson, A. Hanke, Y. Zhang, and S. Levene, J. Comput. Theor. Nanosci. **4**, 1 (2007).
[18] K. Pant, R. L. Karpel, and M. C. Williams, J. Mol. Biol. **327**, 571 (2003).
[19] K. Pant, R. L. Karpel, I. Rouzina, and M. C. Williams, J. Mol. Biol. **336**, 851 (2004); J. Mol. Biol. **349**, 317 (2005).
[20] I. M. Sokolov, R. Metzler, K. Pant, and M. C. Williams, Biophys. J. **89**, 895 (2005).
[21] T. Ambjörnsson and R. Metzler, Phys. Rev. E **72**, 030901(R) (2005).
[22] C. H. Choi, G. Kalosakas, K. Ø. Rasmussen, M. Hiromura, A. R. Bishop, and A. Usheva, Nucleic Acids Res. **32**, 1584 (2004).
[23] S. Ares and G. Kalosakas, Nano Lett. **7**, 307 (2007).
[24] T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, Phys. Rev. Lett. **97**, 128105 (2006).
[25] T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, Biophys. J. **92**, 2674 (2007).
[26] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[27] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
[28] B. S. Alexandrov, L. T. Wille, K. Ø. Rasmussen, A. R. Bishop, and K. B. Blagoev, Phys. Rev. E **74**, 050901 (2006).
[29] A. Campa and A. Giansanti, Phys. Rev. E **58**, 3585 (1998).
[30] M. Peyrard, Nonlinearity **17**, R1 (2004).
[31] A. Hanke and R. Metzler, J. Phys. A **36**, L473 (2003).
[32] A. Bar, Y. Kafri, and D. Mukamel, Phys. Rev. Lett. **98**, 038103 (2007).
[33] H. C. Fogedby and R. Metzler, Phys. Rev. Lett. **98**, 070601 (2007); Phys. Rev. E **76**, 061915 (2007).
[34] D. J. Bicout and E. Kats, Phys. Rev. E **70**, 010902(R) (2004).
[35] T. Ambjörnsson and R. Metzler, J. Phys.: Condens. Matter **17**, S1841 (2005).
[36] S. K. Banik, T. Ambjörnsson, and R. Metzler, Europhys. Lett. **71**, 852 (2005).
[37] T. Hwa, E. Marinari, K. Sneppen, and L.-H. Tang, Proc. Natl. Acad. Sci. U.S.A. **100**, 4411 (2003).
[38] J.-H. Jeon, P. J. Park, and W. Sung, J. Chem. Phys. **125**, 164901 (2006).
[39] T. Ambjörnsson, S. K. Banik, M. A. Lomholt, and R. Metzler, Phys. Rev. E **75**, 021908 (2007).
[40] R. Metzler and T. Ambjörnsson, J. Comput. Theor. Nanosci. **2**, 389 (2005).
[41] T. Ambjörnsson and R. Metzler, J. Phys.: Condens. Matter **17**, S4305 (2005).
[42] D. Poland and H. A. Scheraga, J. Chem. Phys. **45**, 1464 (1966).
[43] M. E. Fisher, J. Chem. Phys. **45**, 1469 (1966).
[44] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
[45] A. Hanke and R. Metzler, Phys. Rev. Lett. **90**, 159801 (2003); Y. Kafri, D. Mukamel, and L. Peliti, *ibid.* **90**, 159802 (2003).
[46] A. Hanke, M. G. Ochoa, and R. Metzler, Phys. Rev. Lett. **100**, 018106 (2008).
[47] R. Blossey and E. Carlon, Phys. Rev. E **68**, 061911 (2003).
[48] G. Bonnet, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. U.S.A. **95**, 8602 (1998); G. Bonnet, S. Tyagi, A. Libchaber, and F. R. Kramer, *ibid.* **96**, 6171 (1999).
[49] O. Krichevsky and G. Bonnet, Rep. Prog. Phys. **65**, 251 (2002).
[50] M. E. Fisher, J. Stat. Phys. **34**, 667 (1984).
[51] A. J. Bray and K. Winkler, J. Phys. A **37**, 5493 (2004).
[52] M. Fixman and J. J. Freire, Biopolymers **16**, 2693 (1977).
[53] Also, $t_L^+(X=-N_L-1,Y)=0$ and $t_R^-(X,Y=N+N_R+1)=0$ for completeness.
[54] E. A. Di Marzio, C. M. Guttman, and J. D. Hoffman, Faraday Discuss. **68**, 210 (1979).
[55] T. Novotný, J. N. Pedersen, M. S. Hansen, T. Ambjörnsson, and R. Metzler, Europhys. Lett. **77**, 48001 (2007).
[56] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1989).
[57] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 2nd ed. (North-Holland, Amsterdam, 1992).
[58] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, Berlin, 1989).
[59] R. F. Pawula, Phys. Rev. **162**, 186 (1967).
[60] J. Marcinkiewicz, Math. Z. **44**, 612 (1939).
[61] The name vicious stems from Ref. 50.
[62] S. Redner, *A Guide to First-Passage Processes* (Cambridge University Press, Cambridge, UK, 2001).
[63] T. Novotný and P. Chvosta, Phys. Rev. E **63**, 012102 (2000).
[64] T. Novotný and R. Metzler (unpublished).
[65] C. Aslangul, J. Phys. A **32**, 3993 (1999).
[66] T. Ambjörnsson and R. J. Silbey, J. Chem. Phys. **129**, 165103 (2008).
[67] M. D. Frank-Kamenetskii, Biopolymers **10**, 2623 (1971).
[68] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii, Nucleic Acids Res. **34**, 564 (2006).
[69] E. Protozanova, P. Yakovchuk, and M. D. Frank-Kamenetskii, J. Mol. Biol. **342**, 775 (2004).
[70] C. Schildkraut and S. Lifson, Biopolymers **3**, 195 (1965).
[71] B. van den Broek, M. A. Lomholt, S.-M. J. Kalisch, R. Metzler, and G. J. L. Wuite, Proc. Natl. Acad. Sci. U.S.A. **105**, 15738 (2008).
[72] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).