DNA breathing dynamics: Analytic results for distribution functions of relevant Brownian functionals

Malay Bandyopadhyay,[1] Shamik Gupta,[2] and Dvira Segal[1]
[1]*Chemical Physics Theory Group, University of Toronto, 80, Saint George Street, Ontario M5S 3H6, Canada*
[2]*Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel*
(Received 3 December 2010; published 9 March 2011)

We investigate DNA breathing dynamics by suggesting and examining several Brownian functionals associated with bubble lifetime and reactivity. Bubble dynamics is described as an overdamped random walk in the number of broken base pairs. The walk takes place on the Poland-Scheraga free-energy landscape. We suggest several probability distribution functions that characterize the breathing process, and adopt the recently studied backward Fokker-Planck method and the path decomposition method as elegant and flexible tools for deriving these distributions. In particular, for a bubble of an initial size x_0, we derive analytical expressions for (i) the distribution $P(t_f|x_0)$ of the first-passage time t_f, characterizing the bubble lifetime, (ii) the distribution $P(A|x_0)$ of the area A until the first-passage time, providing information about the effective reactivity of the bubble to processes within the DNA, (iii) the distribution $P(M)$ of the maximum bubble size M attained before the first-passage time, and (iv) the joint probability distribution $P(M,t_m)$ of the maximum bubble size M and the time t_m of its occurrence before the first-passage time. These distributions are analyzed in the limit of small and large bubble sizes. We supplement our analytical predictions with direct numerical simulations of the related Langevin equation, and obtain a very good agreement in the appropriate limits. The nontrivial scaling behavior of the various quantities analyzed here can, in principle, be explored experimentally.

DOI: 10.1103/PhysRevE.83.031905 PACS number(s): 87.14.gk, 87.10.Mn, 02.50.−r, 05.40.−a

I. INTRODUCTION

The Watson-Crick double-helix structure of DNA derives its stability from the phosphodiester bonds in the single-stranded sugar backbone, and from the hydrogen bonds between complementary base pairs on opposite strands [1,2]. In practice, access to the inside of the double helix, and therefore, the unzipping of a specific region of base pairs is essential for all physiological processes involving DNA, e.g., for replication, transcription, and protein binding [3].

Several mechanisms, such as heating [4], changing the pH of the environment [5], and application of external force [6,7], can lead to unzipping of the double-stranded DNA. This phenomenon is referred to as DNA denaturation. The process occurs progressively, starting with the double strand separating locally into single strands to form loops, or, "bubbles." These bubbles fluctuate in size through stepwise zipping and unzipping of the base pairs at the two zipper forks, where the bubble connects to the double strand. At low temperatures, bubbles once formed eventually close again in time. With the increase of temperature, however, the bubbles grow in size in time to ultimately coalesce with neighboring bubbles and complete the denaturation process. The melting temperature T_m is defined as that at which half of the DNA molecule is denatured, and has typical values $\sim$70–100°C for standard salt solutions [5]. The number of bubbles varies from only a few ones well below T_m up to several hundreds close to T_m.

Breathing dynamics, referring to the dynamics of fluctuating DNA bubbles, has been a topic of intense research for many years [3,4]. It has recently regained interest with the development of new experimental tools that allow for the direct observation of the dynamics of a single DNA molecule [8,9]. Recent experimental studies have further emphasized the key role of DNA breathing dynamics as a determinant of transcription start site location [10,11]. For example, it was demonstrated that an artificial bubble-containing DNA template is transcribed by human RNA polymerase in the absence of any other transcription factors [10]. It was thus concluded that both DNA dynamics and transcription factor binding are necessary conditions for transcription initiation, and that DNA should be treated as an *active* participant in gene expression, rather than as a passive information storage device [11].

On the theoretical side, various methods have been used to study different aspects of the breathing process [12,13], and to investigate the interaction of the DNA with binding proteins: the master equation approach [14,15], a stochastic Gillespie scheme [16], the Fokker-Planck equation approach based on the Poland-Scheraga free-energy function [5,17–19], and stochastic dynamics simulations based on the Peyrard-Bishop-Dauxois (PBD) model for the DNA [10,20,21]. Specifically, the thermally induced denaturation problem has been recently studied by mapping it onto a quantum Coulomb problem [22,23]. These studies have enhanced our understanding of general aspects of both polymer dynamics as well as specific biochemical processes. For recent advances, see Ref. [12].

DNA breathing occurs on a time scale shorter than the equilibration time of the single strands forming the bubbles [9]. Based on this observation, breathing dynamics may be regarded as a random walk in the one-dimensional coordinate x, the number of broken base pairs. Ignoring heterogeneity in the DNA structure, this random walk may be modeled as a noisy overdamped motion at a finite temperature T on the Poland-Scheraga free-energy landscape, $\mathcal{F}(x) \sim \gamma x + ck_BT \ln x$ [13,22]. The parameters $c > 0$ and γ (which can be of either sign) are defined later in the paper. As we show below, this form of the free energy implies a characteristic

scale, $x_{\rm ch}$. For small bubbles ($x < x_{\rm ch}$), the random walk takes place in a potential $\sim \ln x$. In the opposite limit, the potential grows linearly with x, thereby implying a different dynamics. Furthermore, the sign of γ, as given by the system temperature, determines the nature (attractive or repulsive) of the potential. For $\gamma > 0$, which happens at temperatures $T < T_m$, the potential is attractive for all bubble sizes, thereby implying an eventual bubble closure ($x = 0$). On the other hand, above T_m, when $\gamma < 0$, a large bubble ($x > x_{\rm ch}$) evolves under a repulsive linear potential to grow in size toward full denaturation, while a small bubble ($x < x_{\rm ch}$) may still shrink in size to closure under the influence of the attractive $\ln x$ potential.

In this work, following the above picture, we complement previous single-bubble studies by suggesting and analyzing new measures for exploring the DNA breathing process. We focus on several first-passage "Brownian" functionals [24] of the fluctuating bubble, which eventually closes again, and derive their probability distribution functions (pdfs). We separately study the small and large bubble limits, which exhibit different behaviors. The functionals of interest characterize the lifetime of the bubble, the time-integrated bubble size until the first-passage time (the bubble "area"), its maximum size before closure, and the characteristic time for attaining the maximum size. These measures are relevant for estimating the effective reactivity of the bubble, e.g., its efficiency for binding processes.

Another objective of this work is to advocate the use of the recently studied backward Fokker-Planck (BFP) method [25] and the path decomposition (PD) method [26], which builds on the Feynman-Kac formalism [27], for exploring DNA bubble dynamics. These techniques have been extremely useful in studying many aspects of classical Brownian motion, as well as for exploring related problems in computer science and astronomy [25,28,29]. Here, we adopt these elegant methods in the context of DNA breathing dynamics. Using the BFP method, we derive and solve differential equations for the Laplace transforms of various Brownian functionals. This is in contrast to the standard Fokker-Planck treatment, which yields the distribution function to obtain a bubble of a given size at a given time [17–19,21]. Utilizing the PD approach, we can calculate the distribution functions of interest by splitting a representative path of the dynamics into parts, and then considering the weight of each part separately. This is justified by the Markovian property of the dynamics.

In order to gain a qualitative understanding of the DNA breathing process, we separately consider the cases of small and large bubbles. Our analysis reveals interesting scaling laws for the pdfs of various Brownian functionals, which are evidently distinct for small and large bubbles. We further compare our analytical predictions with direct numerical simulations of the corresponding Langevin equation and observe a very good agreement in the appropriate limits.

The paper is organized as follows. In Sec. II, we recall the random-walk model and discuss the distribution functions of interest and their relevance to the DNA dynamics. The BFP and the PD methods are also explained, along with a short description of the numerical technique adopted. In Sec. III, we study the dynamics of small bubbles, and derive the probability distribution functions of several first-passage Brownian functionals. In Sec. IV, the dynamics of large bubbles is examined. We draw our conclusions in Sec. V.

II. MODEL, QUANTITIES OF INTEREST, AND METHODS

A. Model

We follow the Poland-Scheraga approach, and interpret bubbles as occurring owing to free-energy changes to the double-helical ground state [5]. Measuring the size of a bubble by the number of broken base pairs, and denoting this number by the continuous variable $x \geqslant 0$, the Poland-Scheraga free energy is given by [5]

$$\mathcal{F}(x) = \gamma_0 + \gamma x + c k_B T \ln x, \tag{1}$$

where γ_0 is the free-energy barrier to form the initial bubble, while the term γx stands for the free energy required in breaking x base pairs. The entropy loss in forming a closed polymer loop is taken into account by the term $c k_B T \ln x$, where k_B is the Boltzmann constant, T is the temperature, while c is a universal constant determined by the loop configurations [30]. In Eq. (1), a cutoff at $x \sim 1$ is implied. The parameter γ is assumed to have the simple form, $\gamma = \gamma_1(T_m - T)/T_m$, where $\gamma_1 = 4k_B T_r$, with $T_r = 310$ K being the reference temperature.

At finite temperatures, the stochastic dynamics of DNA breathing can be modeled by the overdamped Langevin equation [13,22],

$$\frac{dx}{d\tau} = -D\frac{d\mathcal{F}}{dx} + \xi(\tau). \tag{2}$$

Here, $\xi(\tau)$ is a Gaussian white noise with $\langle \xi(\tau) \rangle = 0$, and $\langle \xi(\tau)\xi(\tau') \rangle = 2Dk_B T\delta(\tau - \tau')$. The kinetic coefficient D has the dimension of $(k_B T)^{-1} s^{-1}$. Using the free energy (1) in Eq. (2), and redefining the time variable, $t \equiv 2Dk_B T\tau$, results in the equation

$$\frac{dx}{dt} = C_2 - \frac{C_1}{x} + \tilde{\xi}(t), \tag{3}$$

where $C_1 = c/2$, $C_2 = \gamma_1(T - T_m)/(2k_B T T_m)$, and

$$\begin{aligned} \langle \tilde{\xi}(t) \rangle &= 0, \\ \langle \tilde{\xi}(t)\tilde{\xi}(t') \rangle &= \delta(t - t'). \end{aligned} \tag{4}$$

Equation (1) implies a characteristic bubble size,

$$x_{\rm ch} = \frac{C_1}{|C_2|}, \tag{5}$$

such that for small bubbles, $x < x_{\rm ch}$, the free energy is dominated by the entropic term $\sim \ln x$. Correspondingly, the Langevin dynamics (3) is essentially governed by the term $-C_1/x$. For large bubbles with $x > x_{\rm ch}$, the base-pair dissociation term $\sim \gamma x$ dominates the free energy, and correspondingly, it is the C_2 term that dictates the Langevin dynamics. For $T < T_m$, the Langevin dynamics occurs in an attractive potential for all bubble sizes, thereby ensuring eventual bubble closure. Above T_m, large bubbles with $x > x_{\rm ch}$ grow in size under a repulsive linear potential to ultimately yield full denaturation, while a small bubble with $x < x_{\rm ch}$ may evolve toward closure under the influence of the attractive $\ln x$ potential. We will utilize below the length scale (5) in

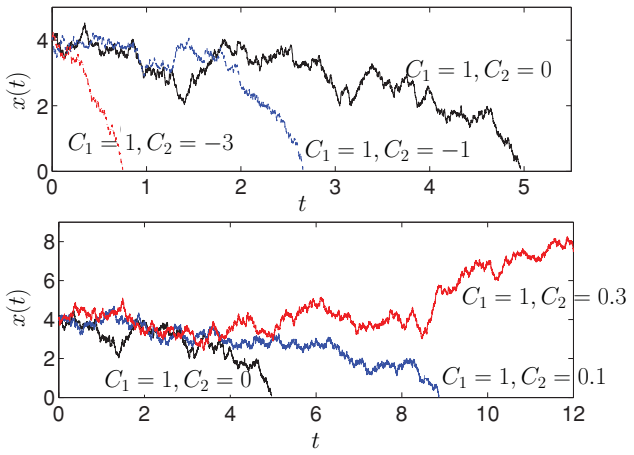

FIG. 1. (Color online) An illustration of several representative paths $x(t)$ following the time evolution of Eq. (3). All the paths begin at $x_0 = 4$. The values of the parameters C_1 and C_2 are marked in the figure. These paths have been generated by using a Brownian simulation (see Sec. II E).

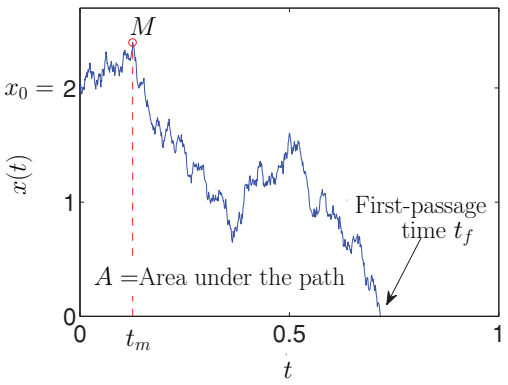

FIG. 2. (Color online) An illustrative path $x(t)$ that begins at x_0 and evolves under Eq. (3). Here, t_f marks the time the path crosses the origin for the first time (corresponding to bubble closure), A is the area enclosed under the path, M and t_m represent, respectively, the maximum value that the path reaches before the first-passage time and the corresponding time of occurrence. This path has been generated by using a Brownian simulation (see Sec. II E) with $C_1 = 1, C_2 = -1$, and $x_0 = 2$.

distinguishing between small and large bubbles. Note that at precisely the melting temperature T_m, when $C_2 = 0$ and the characteristic bubble size diverges, the Langevin dynamics becomes identical to that for small bubbles at all temperatures $T \neq T_m$.

Figure 1 depicts several characteristic paths evolving under Eq. (3) by adopting different values of the parameter C_2 with a fixed C_1. If $|C_2| > C_1$ and $C_2 < 0$ (i.e., $T < T_m$), the bubble closes sufficiently fast in comparison to the case $C_2 = 0$ (top panel). In the opposite limit, taking positive values for C_2, one observes the melting process reflected in the divergence of the first-passage time (bottom panel).

B. Quantities of interest

Our primary focus is on several first-passage Brownian functionals of experimental relevance. We will consider the following quantities and explore their pdfs for small and large bubbles.

(i) *Bubble lifetime.* The first-passage time pdf $P(t_f|x_0)$, i.e., the pdf of the time of closure for bubbles of initial size x_0, provides information about bubble lifetime. A related quantity is the survival probability $C(x_0,t) \equiv 1 - \int_0^t P(x_0|t_f)dt_f$, which can be inferred from experiments by measuring fluorescence correlations of a tagged DNA [8,9].

(ii) *Area under a path.* For the path described by Eq. (3), we define the area under the path before the first-passage time as $A = \int_0^{t_f} x(t')dt'$ (see Fig. 2), and calculate its pdf $P(A|x_0)$. This quantity is of interest because it provides a measure for the effectiveness of breathing-assisted processes, i.e., binding of proteins to the reactive sites of the DNA bases. As an example, consider a process that can take place only inside the double helix, on the single-stranded DNA. Let us assume that this process is facilitated with increasing bubble size, and that it requires a sufficiently long bubble lifetime. While the first-passage time distribution provides information about the average bubble lifetime, it does not contain any hint of the average bubble size before closure. Similarly, $P(M)$, the distribution of the maximum bubble opening before closure provides a measure for the bubble size, but it does not inform us about the corresponding time scale. Thus, we propose the pdf $P(A|x_0)$ of the area A covered until the first-passage time as a useful quantity that provides a measure of bubble reactivity by containing information about both size of the bubble and its characteristic lifetime.

(iii) *Maximum bubble size M.* Another proposed measure for quantifying bubble reactivity is the distribution of the maximum bubble size before the first-passage time, $P(M)$. Consider again a binding process taking place only inside the double helix. Assume next that, owing to geometrical constraints, the process may materialize only when the bubble is large enough. If the time scale of this process is very short, shorter than the average bubble lifetime, a relevant measure for the bubble reactivity is its maximum opening before closure.

Quantities (i), (ii), and (iii) will be calculated below by following the BFP method discussed in Sec. II C.

(iv) *Maximum size M and the corresponding time t_m.* The joint probability distribution function $P(M,t_m)$ will be investigated here by following the PD method, which builds on the Feynman-Kac formalism [25,26]; see Sec. II D. Using this pdf, one can further calculate the distribution function $P(t_m)$ of the time at which the bubble attains its maximum size before closure. This latter pdf is of interest because it provides information about the (average) time of occurrence of the biggest bubble before closure. Processes taking place inside the DNA, facilitated by increased bubble size, will most likely occur around that time.

Note that the joint probability distribution $P(M,t_m)$ and the marginal distribution $P(M)$ are examples of extreme value statistics [31,32]. In our case, these extreme statistics are with respect to a random time interval, namely, the one until the first-passage time.

Figure 2 illustrates a typical path following Eq. (3). The path begins at x_0 and ends at the origin (bubble closure), staying positive in between. The various measures suggested above are also indicated in the figure.

The proposed quantities may be of interest in determining the role of DNA breathing dynamics in gene expression. Experimental data show that the transcriptional start site is often distinguished by large (longer than 10 base pairs) and long-lived transient openings in the double helix [10,11]. Langevin molecular-dynamics simulations have indicated that the most stable DNA openings do not necessarily coincide with the most probable openings and the highest average strand displacement [10]. The quantities suggested here, namely, $P(t_f|x_0)$, $P(A|x_0)$, and $P(M)$, are therefore useful in this context.

It is also of interest to mention other experimentally relevant quantities that have been examined theoretically by using various methods. Deng and Zhu [33] have studied the local denaturation statistics of the thermalized PBD DNA model as a first-passage time problem in the energy of the system. In this work, a backward Kolmogorov equation was used to study quantities such as the mean base-pair opening time. In Ref. [13], DNA zipping rate was calibrated within a Master equation approach, and was found to be in good agreement with experimental data. In Ref. [10], Langevin molecular-dynamics simulations of the PBD model were performed to obtain the probability of occurrence of bubbles of a given size, the average bubble size, and the average lifetime of bubbles of given shape and composition.

C. The backward Fokker-Planck (BFP) method

Following Ref. [25], we recall here how to calculate the statistical properties of a Brownian functional, defined as

$$T = \int_0^{t_f} U(x(\tau))d\tau. \tag{6}$$

Here, $x(\tau)$ is a path representing the motion (3) that starts at x_0 at time $\tau = 0$ and propagates up to $\tau = t_f$, the first-passage time. In the above equation, $U(x(\tau))$ is a specified function of the path, whose choice depends on the quantity of interest. For example, to compute the distribution of the first-passage time t_f, one chooses $U(x(\tau)) = 1$. For the area distribution until the first-passage time, one should consider $U(x(\tau)) = x$. To find the pdf $P(T|x_0)$, noting that the random variable T can be only positive for these choices of $U(x(\tau))$, one considers its Laplace transform

$$Q(x_0,p) = \int_0^\infty P(T|x_0)e^{-pT}dT = \left\langle e^{-p\int_0^{t_f} U(x(\tau))d\tau}\right\rangle, \tag{7}$$

where the angular brackets denote averaging over all paths starting at x_0 at $\tau = 0$ and ending at the first time they cross the origin. For simplicity of notation, in what follows, we suppress the variable p in the function $Q(x_0,p)$. In order to derive a differential equation for $Q(x_0)$, we follow Ref. [25] and split the interval $[0,t_f]$ into two parts. During the first interval $[0,\Delta\tau]$, the path starts from x_0 and propagates up to $x_0 + \Delta x$. In the second interval $[\Delta\tau,t_f]$, the path starts at $x_0 + \Delta x$ and reaches 0 at t_f. Here, $\Delta\tau$ is a fixed, infinitesimally small time interval. We get, to leading order in $\Delta\tau$, $\int_0^{t_f} U(x(\tau))d\tau \approx U(x_0)\Delta\tau + \int_{\Delta\tau}^{t_f} U(x)d\tau$, and hence, from Eq. (7),

$$\begin{aligned} Q(x_0) &\approx e^{-pU(x_0)\Delta\tau}\langle Q(x_0+\Delta x)\rangle_{\Delta x} \\ &\approx (1 - pU(x_0)\Delta\tau)\langle Q(x_0+\Delta x)\rangle_{\Delta x}. \end{aligned} \tag{8}$$

Now, the average denoted by the angular brackets is performed over all realizations of Δx. The dynamical Eq. (3) gives $\Delta x = F(x_0)\Delta\tau + \tilde{\xi}(0)\Delta\tau$, with $F(x_0) = C_2 - C_1/x_0$. Substituting for Δx in Eq. (8), expanding $Q(x_0+\Delta x)$ in powers of $\Delta\tau$, and averaging over the noise by using $\langle\tilde{\xi}(0)\rangle = 0$ and $\langle\tilde{\xi}^2(0)\rangle = 1/\Delta\tau$ for small $\Delta\tau$, one obtains, to lowest order in $\Delta\tau$, the ordinary differential equation,

$$\frac{1}{2}\frac{d^2Q(x_0)}{dx_0^2} + \left(C_2 - \frac{C_1}{x_0}\right)\frac{dQ(x_0)}{dx_0} - pU(x_0)Q(x_0) = 0. \tag{9}$$

Boundary conditions. The above equation is valid for $x_0 \in [0,\infty]$ with the following boundary conditions: (i) For an infinitesimally small bubble, $x_0 \to 0$, the first-passage time vanishes, $t_f \to 0$, so that $Q(x_0 = 0) = 1$. (ii) If the bubble is initially large, $x_0 \to \infty$, the first-passage time diverges, hence, $Q(x_0 \to \infty) = 0$.

We emphasize that the differential Eq. (9), referred to as the BFP equation [25], directly provides us with the Laplace-transformed pdfs of various quantities that are determined by the choice of $U(x)$. In contrast, the standard Fokker-Planck method adopted in Refs. [17–19,21] yields the distribution function $P(x,t)$ to obtain a loop of size x at time t. Thus, these two approaches are distinct, providing complementary information.

D. The path decomposition (PD) method

The principle of this technique is simple: Because the motion in Eq. (3) is Markovian, a typical path can be split into, e.g., two parts. Then, the weight of the whole path is the product of the weights of the two split parts [26].

The above idea allows us to calculate the joint probability distribution $P(M,t_m)$ of the maximum bubble size M and the time t_m at which this maximum occurs before closure, given that the initial size of the bubble is fixed at $x_0 \in [0,M]$. By integrating over M, one can further obtain the marginal distribution $P(t_m)$. We compute $P(M,t_m)$ by splitting a typical path into two parts, before and after t_m, with the respective weights W_L and W_R, so that the weight W of the whole path is

$$W = W_L \times W_R. \tag{10}$$

On the left-hand side of t_m, the path propagates from x_0 at $t = 0$ to $M - \epsilon$ at $t = t_m$, without ever attaining the value 0 or M during the interval $[0,t_m]$ [34]. The weight W_L can be determined by using a path-integral treatment based on the Feynman-Kac formalism, as we explain below. On the right-hand side of t_m, the path starts from $M - \epsilon$ at $t = t_m$ and ends at the origin at t_f (with $t_f \geqslant t_m$), without crossing either the level M or the level 0 in between. At the end of the calculation, one needs to take the limit $\epsilon \to 0$.

The calculation of W_R will be explained in Secs. III D and IV D. We explain here in some detail the calculation of W_L. Because the white noise in Eq. (3) is Gaussian, the probability of a path is given by

$$P[\{x(\tau)\}] \propto \exp\left[-\frac{1}{2}\int_0^t d\tau\left(\frac{dx}{d\tau} + \frac{C_1}{x} - C_2\right)^2\right]. \tag{11}$$

The weight W_L is then given as a sum over contributions from all possible paths,

$$\begin{aligned} W_L &\propto \int_{x(0)=x_0}^{x(t_m)=M-\epsilon} \mathcal{D}x(\tau) \\ &\quad \times \exp\left[-\frac{1}{2}\int_0^{t_m} d\tau \left(\frac{dx}{d\tau} + \frac{C_1}{x} - C_2\right)^2 \right] \\ &\quad \times \prod_{\tau=0}^{t_m} \theta[x(\tau)] \prod_{\tau=0}^{t_m} \theta[M - x(\tau)] \\ &= \left(\frac{x_0}{M-\epsilon}\right)^{C_1} e^{C_2(M-\epsilon-x_0)} \int_{x(0)=x_0}^{x(t_m)=M-\epsilon} \mathcal{D}x(\tau) \\ &\quad \times \exp\left\{ -\int_0^{t_m} d\tau \left[\frac{1}{2}\left(\frac{dx}{d\tau}\right)^2 + \frac{1}{2}\left(\frac{C_1}{x} - C_2\right)^2 \right] \right\} \\ &\quad \times \prod_{\tau=0}^{t_m} \theta[x(\tau)] \prod_{\tau=0}^{t_m} \theta[M - x(\tau)]. \end{aligned} \tag{12}$$

In the above equations, the terms $\prod_{\tau=0}^{t_m}\theta[x(\tau)]$ and $\prod_{\tau=0}^{t_m}\theta[M-x(\tau)]$ enforce the requirements that the path does not cross either the level 0 or the level M for times between 0 and t_m. Following Feynman-Kac [27], the path integral in Eq. (12) is identified with the propagator $\langle M-\epsilon|e^{-\hat{H}t_m}|x_0\rangle$, corresponding to the quantum Hamiltonian $\hat{H}$ of a single particle of unit mass,

$$\hat{H} = -\frac{1}{2}\frac{d^2}{dx^2} + V(x), \tag{13}$$

with $\hbar = 1$. The potential energy $V(x)$ is given by

$$V(x) = \begin{cases} \frac{1}{2}\left(\frac{C_1}{x} - C_2\right)^2 & \text{if} \quad 0 < x < M, \\ \infty & \text{if } x = 0 \quad \text{or} \quad x = M. \end{cases} \tag{14}$$

Note that the infinite potential energy at $x = 0$ and at $x = M$ enforces the requirement that the path never crosses either the level 0 or the level M. Finally, we get

$$\begin{aligned} W_L &\propto \left(\frac{x_0}{M-\epsilon}\right)^{C_1} e^{C_2(M-\epsilon-x_0)} \langle M-\epsilon|e^{-\hat{H}t_m}|x_0\rangle \\ &= \left(\frac{x_0}{M-\epsilon}\right)^{C_1} e^{C_2(M-\epsilon-x_0)} \sum_{p=1}^{\infty} e^{-E_p t_m} \psi_p(M-\epsilon)\psi_p(x_0), \end{aligned} \tag{15}$$

where $\psi_p(x)$ and E_p are the eigenfunctions and eigenenergies, respectively, of the Hamiltonian $\hat{H}$ in Eq. (13). As a result of the infinite potential barrier at $x = 0$ and at $x = M$, the eigenfunctions satisfy $\psi_p(x = 0, M) = 0$.

E. Simulations

The statistical properties of Brownian functionals studied here can be numerically obtained by integrating the overdamped Langevin equation (3). Using a second-order stochastic Runge-Kutta algorithm [35], we update the trajectory by following the rule,

$$x(\Delta t) = x_0 + \tfrac{1}{2}\left[F(x_0) + F(x_0 + F(x_0)\Delta t + \Gamma_0)\right]\Delta t + \Gamma_0, \tag{16}$$

where $F(x) = C_2 - C_1/x$. Here, Γ_0 is a random number sampled from a Gaussian distribution with zero mean and width given by $\langle \Gamma_0^2 \rangle = \Delta t$. For all simulations presented in this work, we take $\Delta t = 10^{-3}$, unless stated otherwise. We generate a large set of paths, all starting at a particular x_0 and ending close to the origin (within a preassigned numerical tolerance value). Averaging over an ensemble, we generate various pdfs that we compare with our analytical results.

III. SMALL BUBBLE DYNAMICS

We begin our analysis by considering small bubbles, $x < x_{\rm ch}$, at all temperatures, $T \neq T_m$. The analysis is also valid for bubbles of all sizes at precisely the melting temperature T_m. In these cases, the nonlinear entropic term in the free energy (1) dictates the dynamics, resulting in the Langevin equation,

$$\frac{dx}{dt} = -\frac{C_1}{x} + \tilde{\xi}, \tag{17}$$

where $\langle \tilde{\xi}(t) \rangle = 0$ and $\langle \tilde{\xi}(t)\tilde{\xi}(t') \rangle = \delta(t - t')$. Equation (17) defines a Bessel process. The pdfs $P(t_f|x_0)$ and $P(A|x_0)$ are obtained by the BFP method, where the differential equation that needs to be solved is given by Eq. (9) with $C_2 = 0$,

$$\frac{1}{2}\frac{d^2Q(x_0)}{dx_0^2} - \frac{C_1}{x_0}\frac{dQ(x_0)}{dx_0} - pU(x_0)Q(x_0) = 0. \tag{18}$$

The boundary conditions are (i) $Q(x_0 \to \infty) = 0$ and (ii) $Q(x_0 \to 0) = 1$. We also derive analytical results for $P(M)$ and $P(M,t_m)$, which, as explained above, are examples of extreme statistics with respect to the first-passage time, i.e., with respect to a time interval that itself is random. We note that a recent study has adopted the real-space renormalization group technique in looking at extreme statistics of the Bessel process over a fixed interval of time [36].

A. First-passage time distribution: $P(t_f|x_0)$

We compute the distribution of t_f, the time at which the bubble closes for the first time, assuming its initial size is fixed at x_0, by substituting $U(x_0) = 1$ in Eq. (18),

$$\frac{1}{2}\frac{d^2Q(x_0)}{dx_0^2} - \frac{C_1}{x_0}\frac{dQ(x_0)}{dx_0} - pQ(x_0) = 0. \tag{19}$$

The general solution of Eq. (19) is [37]

$$Q(x_0) = x_0^{\alpha}[AI_{\alpha}(\sqrt{2p}x_0) + BK_{\alpha}(\sqrt{2p}x_0)]. \tag{20}$$

Here, $I_\alpha(x)$ and $K_\alpha(x)$ are the modified Bessel functions of the first and second kind, respectively. Also, $\alpha = C_1 + 1/2$, and A and B are arbitrary constants to be determined from the boundary conditions. Because for large x, $I_\alpha(x) \sim e^x/\sqrt{2\pi x}$ and $K_\alpha(x) \sim \sqrt{\pi/2x}e^{-x}$ [38], in order to satisfy the condition $Q(x_0 \to \infty) = 0$, we must have $A = 0$. To satisfy the condition $Q(x_0 \to 0) = 1$, we note that as $x \to 0$, $K_\alpha(x) \approx \Gamma(\alpha)2^{\alpha-1}/x^\alpha$ for $\alpha > 0$ [38], which gives $B = (\sqrt{2p})^{C_1+1/2}/[\Gamma(C_1 + 1/2)2^{C_1-1/2}]$. Following these considerations, we get the particular solution,

$$Q(x_0) = x_0^{C_1+1/2}\frac{(\sqrt{2p})^{C_1+1/2}}{\Gamma(C_1 + 1/2)2^{C_1-1/2}}K_{C_1+1/2}(\sqrt{2p}x_0). \tag{21}$$

On taking an inverse Laplace transform, we get [39]

$$P(t_f|x_0) = \frac{x_0^{2C_1+1}}{\Gamma(C_1+1/2)2^{C_1+1/2}} t_f^{-C_1-3/2} e^{-x_0^2/2t_f}, \quad (22)$$

as obtained earlier in Refs. [22] and [40].

Next, we compare the analytical prediction (22) with numerical simulations under the *full* bubble potential, including the C_2 contribution, in order to explore the regime of validity of the above result (see the top panel of Fig. 3). The sampled trajectories all begin at $x_0 = 2$. When $x_{\rm ch} = C_1/|C_2| \gtrsim 10 > x_0$, $C_2 < 0$, we observe a good agreement between numerics and analytical results. For $x_0 \sim x_{\rm ch}$, deviations occur because then the contribution of C_2 cannot be neglected (inset).

Besides the distribution (22), other related quantities of experimental relevance are the moments, $\langle t_f^k \rangle$, obtained from Eq. (22) as

$$\langle t_f^k \rangle = \frac{x_0^{2k}}{2^k} \frac{\Gamma(C_1 - k + 1/2)}{\Gamma(C_1 + 1/2)} \quad \text{for } k > C_1 + 1/2, \quad (23)$$

while $\langle t_f^k \rangle$ diverges for $k < C_1 + 1/2$. Another important quantity is the persistence, or survival probability of the bubble, defined as

$$C(x_0,t) \equiv 1 - \int_0^t P(t_f|x_0) dt_f, \quad (24)$$

where $\int_0^t P(t_f|x_0)dt_f$ sums up the probabilities of all events where the bubble closes in time t. This quantity can be resolved in experiments by measuring fluorescence correlations of a tagged bubble [8,9]. On plugging Eq. (22) into Eq. (24), we get

$$C(x_0,t) = 1 - \frac{\Gamma\left(C_1 + 1/2, x_0^2/2t\right)}{\Gamma(C_1 + 1/2)}, \quad (25)$$

where $\Gamma(s,x) = \int_x^\infty t^{s-1} \exp(-t) dt$ is the upper incomplete gamma function. This result agrees with that reported in Ref. [22]. It is easy to derive the following asymptotic behaviors of $C(x_0,t)$: In the limit $t \to \infty$, one has [22]

$$C(x_0,t) \approx \frac{\left(x_0^2\right)^{C_1+1/2}}{(C_1 + 1/2)\Gamma(C_1 + 1/2)} t^{-C_1-1/2}, \quad (26)$$

while, in the limit $t \to 0$, one has

$$C(x_0,t) \approx 1 - \frac{\left(x_0^2/2\right)^{C_1-1/2}}{\Gamma(C_1 + 1/2)} t^{1/2-C_1} e^{-x_0^2/2t}. \quad (27)$$

B. Distribution of the area until the first-passage time: $P(A|x_0)$

The area $A = \int_0^{t_f} x(t')dt'$ under the random motion (17) can tell us about the readiness of the bubble to react. Here, the motion starts at x_0 and continues in time until the first-passage time. Note that the quantity A is not a geometric area, rather its units are length $\times$ time. To compute the related pdf, we substitute $U(x_0) = x_0$ in Eq. (18),

$$\frac{1}{2}\frac{d^2Q(x_0)}{dx_0^2} - \frac{C_1}{x_0}\frac{dQ(x_0)}{dx_0} - px_0Q(x_0) = 0, \quad (28)$$

which has the general solution [41]

$$Q(x_0) = x_0^{C_1+1/2}[A_1 J_\nu(iz) + A_2 J_{-\nu}(iz)]. \quad (29)$$

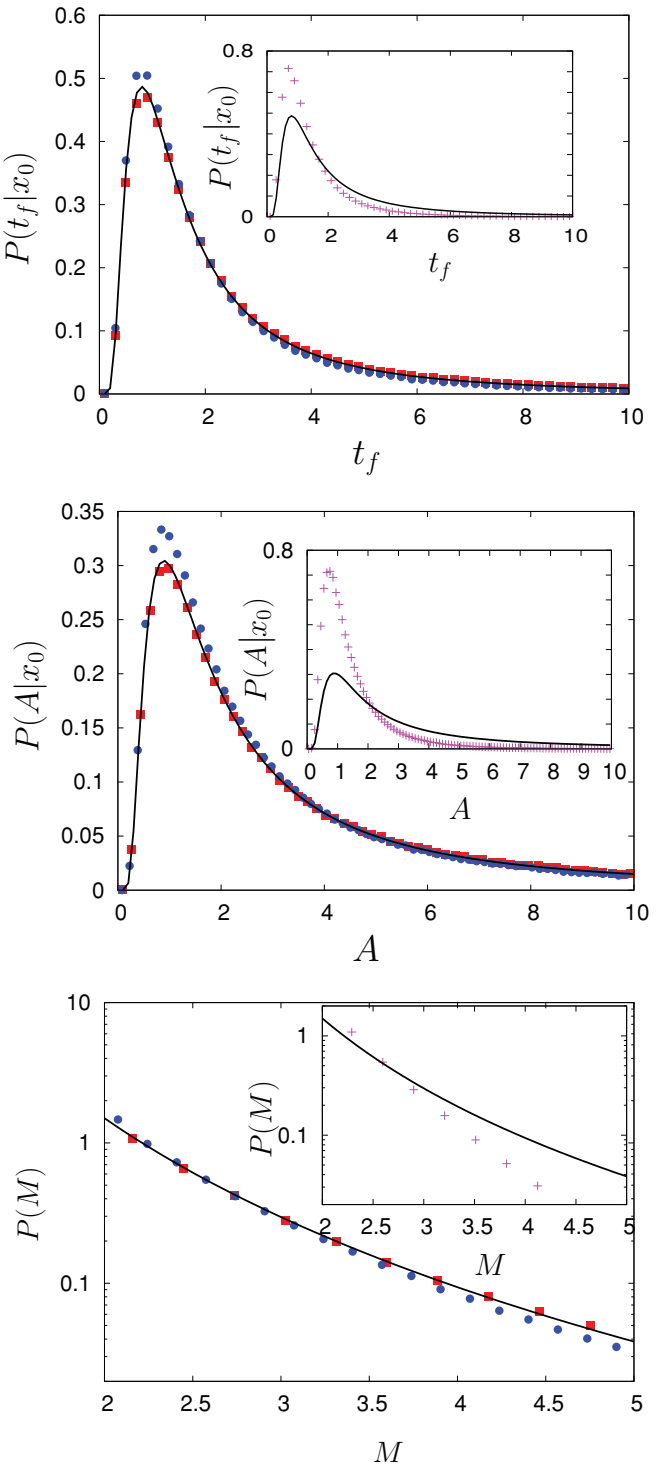

FIG. 3. (Color online) Numerical simulations below the melting temperature, $T < T_m$, for the pdf of the first-passage time (top), the pdf of the area until the first-passage time (center), and the pdf of the maximum size until the first-passage time (bottom). The parameters are $C_1 = 1$ and $C_2 = 0$ (red squares), $C_1 = 1$ and $C_2 = -0.1$ (blue dots). The initial bubble size is $x_0 = 2$ for all the cases. The analytic results, (22), (32), and (36), appear in black continuous lines. The three insets compare the analytic small bubble results (black continuous line) with numerical simulations using $C_1 = 1$ and $C_2 = -0.5$ (purple crosses).

Here, $z = (2/3)\sqrt{2p}x_0^{3/2}$, $\nu = (2C_1+1)/3$, $J_\nu(x)$ is the Bessel function of the first kind, and A_1 and A_2 are arbitrary constants. Using $J_\nu(x) = i^\nu I_\nu(-ix)$, where $I_\nu(x)$ is the modified Bessel function of the first kind [38], gives

$$Q(x_0) = x_0^{C_1+1/2}[B_1 I_\nu(z) + B_2 I_{-\nu}(z)], \tag{30}$$

where B_1 and B_2 are arbitrary constants. Because for large x, $I_{\pm\nu}(x) \sim e^x/\sqrt{2\pi x}$ [38], to satisfy the boundary condition $Q(x_0 \to \infty) = 0$, we must have $B_1 = -B_2$. To satisfy $Q(x_0 \to 0) = 1$, we note that as $x_0 \to 0$, one has $I_\nu(x) \approx (x/2)^\nu/\Gamma(\nu+1)$ [38], which yields $B_2 = \Gamma(1-\nu)(\sqrt{2p}/3)^\nu$. We thus get

$$\begin{aligned} Q(x_0) &= \left(\frac{z}{2}\right)^\nu \Gamma(1-\nu)[I_{-\nu}(z) - I_\nu(z)] \\ &= \left(\frac{z}{2}\right)^\nu \frac{2}{\Gamma(\nu)} K_\nu(z), \end{aligned} \tag{31}$$

where, in obtaining the last relation, we have used the identities, $K_\nu(x) = (\pi/2)[I_{-\nu}(x) - I_\nu(x)]/\sin(\nu\pi)$ and $\Gamma(\nu)\Gamma(1-\nu) = \pi/\sin(\nu\pi)$ [38]. On taking an inverse Laplace transform of (31), we obtain the desired pdf [39]:

$$P(A|x_0) = \frac{2^{(2C_1+1)/3} x_0^{2C_1+1}}{3^{(4C_1+2)/3}\Gamma[(2C_1+1)/3]} \frac{\exp\left(-2x_0^3/9A\right)}{A^{(2C_1+4)/3}}. \tag{32}$$

This expression reproduces well the numerical results obtained by simulating the Langevin equation (3) under the *full* potential with $x_0 < x_{\rm ch}$ and $C_2 < 0$ (see the middle panel of Fig. 3). The inset demonstrates an expected disagreement for larger bubbles with $x_0 \sim x_{\rm ch}$.

C. Distribution of the maximum before the first-passage time: $P(M)$

How large can the bubble grow before it closes, assuming an initial opening of x_0? This question is of interest in estimating the efficiency of processes that can occur inside big loops only. The relevant measure is provided by $P(M)$, the pdf of the maximum bubble size M before its first closure, given that $x_0 \in [0,M]$. We obtain this pdf by following closely the procedure of Refs. [26] and [42]. We first define a related function. Let $q(x_0)$ be the probability that the motion described by Eq. (17) exits the interval $[0,M]$ for the first time through the origin. Thus, $q(x_0)$ is the cumulative probability that the maximum before the first-passage time is $\leqslant M$. It is evident that this function satisfies two boundary conditions: (i) $q(0) = 1$ and (ii) $q(M) = 0$. Denoting by $\phi_{\Delta\tau}(\Delta x)$ the distribution function of a small displacement Δx in time $\Delta\tau \to 0$, it follows from the Markovian property of the dynamics (17) that

$$q(x_0) = \int q(x_0 + \Delta x)\phi_{\Delta\tau}(\Delta x)d(\Delta x). \tag{33}$$

On Taylor expanding $q(x_0+\Delta x)$ and averaging over $\Delta x = -(C_1/x_0)\Delta\tau + \tilde{\xi}(0)\Delta\tau$, using $\langle\tilde{\xi}(0)\rangle = 0$, $\langle\tilde{\xi}^2(0)\rangle = 1/\Delta\tau$, we get, to leading order in $\Delta\tau$, the equation $\{(1/2)[d^2q(x_0)/dx_0^2] - (C_1/x_0)[dq(x_0)/dx_0]\}\Delta\tau = 0$. For arbitrary $\Delta\tau$, we obtain

$$\frac{1}{2}\frac{d^2q(x_0)}{dx_0^2} - \frac{C_1}{x_0}\frac{dq(x_0)}{dx_0} = 0. \tag{34}$$

Solving this equation with the above-mentioned boundary conditions, we get

$$q(x_0) = 1 - \left(\frac{x_0}{M}\right)^{2C_1+1}. \tag{35}$$

The pdf of interest is obtained by differentiating $q(x_0)$ with respect to M,

$$P(M) = \frac{(2C_1+1)x_0^{2C_1+1}}{M^{2C_1+2}}, \quad M \geqslant x_0. \tag{36}$$

In Fig. 3 (bottom), we compare this result with numerical simulations using the full potential. We observe good agreement for $x_0 < x_{\rm ch}$, $C_2 < 0$, and an expected disagreement for $x_0 \sim x_{\rm ch}$ (inset).

D. Joint pdf of the maximum M and the corresponding time t_m before the first-passage time: $P(M,t_m)$

To compute the joint probability distribution of the maximum bubble size M and the time t_m at which the maximum occurs before closure, we adopt the PD method (see Sec. II D and Ref. [26]). We split a path evolving under (17) into two parts, before and after the time t_m, with respective weights W_L and W_R. Owing to the Markovian property, the weight W of the whole path is given by the product of the weights for the two split parts, $W = W_L \times W_R$. The weight W_R can be obtained from Eq. (35). Recall that W_R is the weight of a path that starts at $M-\epsilon$ at time $t = t_m$ and exits the interval $[0,M]$ for the first time through the origin. On the other hand, $q(x_0)$ in Eq. (35) is the probability for a path starting at $x_0 \in [0,M]$ to exit the interval for the first time through the origin. We thus deduce that $W_R = q(M-\epsilon)$, or

$$W_R = 1 - \frac{(M-\epsilon)^{(2C_1+1)}}{M^{2C_1+1}} = \frac{(2C_1+1)\epsilon}{M} + O(\epsilon^2). \tag{37}$$

The second equality is derived by assuming ϵ to be infinitesimal. The weight W_L is obtained from Eq. (15) by substituting $C_2 = 0$,

$$W_L \propto \left(\frac{x_0}{M-\epsilon}\right)^{C_1} \sum_{p=1}^{\infty} e^{-E_p t_m}\psi_p(M-\epsilon)\psi_p(x_0). \tag{38}$$

Here, ψ_p and E_p are the solutions of the eigenequation,

$$\left[-\frac{1}{2}\frac{d^2}{dx^2} + \frac{C_1^2}{2x^2}\right]\psi = E\psi, \quad 0 < x < M, \tag{39}$$

subject to the condition $\psi(x = 0,M) = 0$. The general solution of this equation is

$$\psi_p(x) = A\sqrt{x}J_\alpha(\sqrt{2E_p}x) + B\sqrt{x}Y_\alpha(\sqrt{2E_p}x), \tag{40}$$

where $J_\alpha(x)$ and $Y_\alpha(x)$ are the Bessel functions of order α of the first and second kind, respectively, and $\alpha = \frac{1}{2}\sqrt{1+4C_1^2}$ [37]. Note that C_1 is real, thus $\alpha > 0$. Because for $x \to 0$, $Y_\alpha(x) \approx -[\Gamma(\alpha)/\pi](2/x)^\alpha$ [38], we demand that $B = 0$ for satisfying $\psi(x = 0) = 0$. The other boundary condition results in the discrete eigenvalues E_p such that $\sqrt{2E_p}M = u_{\alpha p}$, where $u_{\alpha p}$ denotes the pth zero of $J_\alpha(x)$. The constant A is determined by requiring $\psi_p(x)$ to be normalized.

On using the identity, $\int_0^a d\rho\rho J_\alpha(u_{\alpha p}\rho/a)J_\alpha(u_{\alpha q}\rho/a) = \delta_{p,q}(a^2/2)[J_{\alpha+1}(u_{\alpha p})]^2$ [38], we finally get

$$\psi_p(x) = \frac{\sqrt{2x}}{M|J_{\alpha+1}(u_{\alpha p})|} J_\alpha\left(\frac{u_{\alpha p}x}{M}\right), \tag{41}$$

and the probability

$$W_L \propto \left(\frac{x_0}{M-\epsilon}\right)^{C_1} \frac{2\sqrt{(M-\epsilon)x_0}}{M^2} \times \sum_{p=1}^{\infty} \frac{e^{-u_{\alpha p}^2 t_m/(2M^2)}}{[J_{\alpha+1}(u_{\alpha p})]^2} J_\alpha\left(\frac{u_{\alpha p}(M-\epsilon)}{M}\right) J_\alpha\left(\frac{u_{\alpha p}x_0}{M}\right). \tag{42}$$

Next, we evaluate W_L to leading order in ϵ by Taylor expanding $J_\alpha(\frac{u_{\alpha p}(M-\epsilon)}{M})$ and also using the result $J'_\alpha(u_{\alpha p}) = -J_{\alpha+1}(u_{\alpha p})$, which follows from the following identity: $J'_\alpha(z) = \frac{\alpha}{z}J_\alpha(z) - J_{\alpha+1}(z)$ [37]. We finally get

$$W_L \propto \epsilon \frac{2x_0^{C_1+1/2}}{M^{C_1+5/2}} \sum_{p=1}^{\infty} u_{\alpha p} \frac{e^{-u_{\alpha p}^2 t_m/(2M^2)}}{J_{\alpha+1}(u_{\alpha p})} J_\alpha\left(\frac{u_{\alpha p}x_0}{M}\right) + O(\epsilon^2). \tag{43}$$

The probability $P(M,t_m;\epsilon)$ of the whole path is the product of (43) and (38) with a normalization constant $B(\epsilon)$, which is determined by requiring that $\lim_{\epsilon\to 0}\int_0^\infty P(M,t_m;\epsilon)dt_m \to P(M)$, where $P(M)$ is given by Eq. (36),

$$P(M,t_m;\epsilon) = B(\epsilon)W_L W_R. \tag{44}$$

Using the identity $\sum_{p=1}^{\infty} J_\alpha(u_{\alpha p}x_0/M)/[u_{\alpha p}J_{\alpha+1}(u_{\alpha p})] = x_0^\alpha/(2M^\alpha)$ for $0 \leqslant x_0 < M$ [43], we get $B(\epsilon) = \frac{1}{2\epsilon^2}(\frac{x_0}{M})^{C-C_1-1/2}$, where $C = 2C_1 + 1 - \sqrt{1+4C_1^2}/2$. Substituting for $B(\epsilon)$ in Eq. (44) and taking the limit $\epsilon \to 0$, we get the desired probability,

$$P(M,t_m) = (2C_1+1)\frac{x_0^C}{M^{C+3}} \times \sum_{p=1}^{\infty} u_{\alpha p} \frac{e^{-u_{\alpha p}^2 t_m/(2M^2)}}{J_{\alpha+1}(u_{\alpha p})} J_\alpha\left(\frac{u_{\alpha p}x_0}{M}\right). \tag{45}$$

For the free Brownian motion, taking $C_1 = 0$, this result reduces to that derived in Ref. [26].

IV. LARGE BUBBLE DYNAMICS

We study here the dynamics of large bubbles of size $x > x_{\rm ch}$ [see Eq. (5)]. In this limit, one can neglect the term $-C_1/x$ in the Langevin equation (3), and study the dynamics dictated by

$$\frac{dx}{dt} = C_2 + \tilde{\xi}(t). \tag{46}$$

This equation describes a one-dimensional random walk, $x(t)$, in the presence of a constant drift, C_2. The probability distribution to find a bubble of size x at time t is a shifted Gaussian $P(x,t) \sim e^{-(x-x_0-C_2t)^2/2t}$, where x_0 is the initial size of the bubble. Below the melting temperature ($C_2 < 0$), the bubble shrinks and eventually disappears in time. Above T_m, when $C_2 > 0$, the bubble grows in size as time passes. For large $C_2 > 0$, the first-passage time diverges. Thus, our analysis below of various first-passage functionals is valid for large bubbles at temperatures below T_m as well as above T_m, in the latter case while conditioned on a finite t_f.

For obtaining the pdfs, $P(t_f|x_0)$ and $P(A|x_0)$, we adopt the BFP method. The relevant differential equation is obtained from Eq. (9) by substituting $C_1 = 0$,

$$\frac{1}{2}\frac{d^2Q(x_0)}{dx_0^2} + C_2\frac{dQ(x_0)}{dx_0} - pU(x_0)Q(x_0) = 0, \tag{47}$$

with the boundary conditions (i) $Q(x_0 \to \infty) = 0$ and (ii) $Q(x_0 \to 0) = 1$. We also derive analytical results for $P(M)$ and $P(M,t_m)$.

A. First-passage time distribution: $P(t_f|x_0)$

This distribution has already been investigated in Ref. [42] in a different context. We thus omit the details of the calculation, but include the results for the sake of completeness of our presentation. The procedure involves solving Eq. (47) with $U(x_0) = 1$ under the boundary conditions, then taking inverse Laplace transform of the solution to yield

$$P(t_f|x_0) = \frac{1}{\sqrt{2\pi}}\frac{x_0}{t_f^{3/2}}\exp\left[-\frac{(x_0+C_2t_f)^2}{2t_f}\right]. \tag{48}$$

The moments $\langle t_f^k\rangle$ may be obtained by using the identity, $\int_0^\infty x^{\nu-1}e^{-\beta/x-\gamma x}dx = 2(\beta/\gamma)^{\nu/2}K_\nu(2\sqrt{\beta\gamma})$ for $\mathrm{Re}(\beta) > 0$ and $\mathrm{Re}(\gamma) > 0$, where $K_\nu(x)$ is the modified Bessel function of the second kind [37]. One gets

$$\langle t_f^k\rangle = \sqrt{\frac{2}{\pi}}\left(\frac{x_0}{|C_2|}\right)^k (x_0|C_2|)^{1/2}e^{-C_2x_0}K_{k-1/2}(|C_2|x_0). \tag{49}$$

Noting that $K_{1/2}(x) = \sqrt{\pi/(2x)}e^{-x}$, the mean first-passage time is given by $\langle t_f\rangle = x_0e^{-2C_2x_0}/C_2$ for $C_2 > 0$, and by $\langle t_f\rangle = x_0/|C_2|$ for $C_2 < 0$. The survival probability, defined in Eq. (24), is given by

$$C(x_0,t) = 1 - \frac{x_0}{\sqrt{2\pi}}e^{-C_2x_0}\int_0^t t_f^{-3/2}e^{-x_0^2/(2t_f)-C_2^2t_f/2}dt_f. \tag{50}$$

For $C_2 < 0$ ($T < T_m$) and for large t, one has $C(x_0,t) \approx 1-(C_2x_0)/(2\sqrt{\pi})e^{-C_2x_0-x_0^2/(2t)}[\Gamma(-1/2) - \Gamma(-1/2,C_2^2t/2)]$. Using the result that $\Gamma(s,x) \to x^{s-1}e^{-x}$ as $x \to \infty$ [38], we get $C(x_0,t) \sim \sqrt{2/\pi}(x_0/C_2^2)t^{-3/2}\exp[-(x_0-|C_2|t)^2/(2t)]$.

Figure 4 compares the analytic result for $P(t_f|x_0)$ with numerical simulations obtained by considering the full potential. Because, for $x_0 > x_{\rm ch}$ with $C_2 > 0$, many trajectories have diverging t_f, we performed here numerical simulations only below the melting temperature, so that $C_2 < 0$. In this case, with increasing $|C_2|$, the effective drift velocity toward bubble closure increases and bubbles quickly disappear in time. The top panel demonstrates that the analytical prediction (48) agrees with simulation results for $x_0 > x_{\rm ch}; C_2 < 0$. Figure 5 further includes results from numerical simulations by considering the full bubble potential, displaying the behaviors both below and above the melting temperature. Upon increasing C_2 from negative values (i.e., for $T < T_m$) to positive values (i.e., for $T > T_m$), one notes that (i) the center of the pdf $P(t_f|x_0)$ is displaced to longer times, and that (ii) the bubble

lifetime is significantly enhanced (see Fig. 5, top panel). This corroborates with the physical picture that with increasing temperature, a bubble takes a longer time to disappear.

B. Distribution of the area until the first-passage time: $P(A|x_0)$

This quantity can be obtained by solving Eq. (47) with $U(x_0) = x_0$ with the appropriate boundary conditions, then deriving the inverse Laplace transform of the solution (see Refs. [42] and [44] for details). In particular, one obtains the following two limiting behaviors of the distribution $P(A|x_0)$: For $A \to \infty$, one has

$$P(A|x_0) \approx \frac{e^{-C_2 x_0}\sinh(|C_2|x_0)}{\sqrt{\pi}}\left(\frac{2}{3}\right)^{1/4}\left(\frac{|C_2|}{A}\right)^{3/4} \times \exp\left[-\left(\frac{8}{3}\right)^{1/2}|C_2|^{3/2}A^{1/2}\right]. \quad (51)$$

In the opposite $A \to 0$ limit, one gets [44]

$$P(A|x_0) \approx \frac{2^{1/3}}{3^{2/3}\Gamma(1/3)}\frac{x_0 e^{-C_2 x_0}}{A^{4/3}}e^{-2x_0^3/9A}. \quad (52)$$

Note the distinct asymptotic forms in Eqs. (51) and (52). While the latter demonstrates a behavior similar to that observed in the small bubble dynamics [cf. Eq. (32)], the former predicts a different scaling behavior.

Results from numerical simulations are displayed in Figs. 4 and 5. For a fixed value of $C_2 < 0$, on increasing C_1, Fig. 4 shows the narrowing of $P(A|x_0)$ and the displacement of its center toward smaller A values, thereby reflecting the increased importance of the entropy term in the free-energy function. On fixing C_1 and on increasing C_2 from negative to (small) positive values, the area pdf develops an increasing contribution at large A values, thereby hinting at the onset of large bubbles and long first-passage times when $T > T_m$.

C. Distribution of the maximum before the first-passage time: $P(M)$

The procedure here proceeds as in Sec. III C. One first finds the cumulative probability $q(x_0)$, which satisfies [26,42]

$$\frac{1}{2}\frac{d^2q(x_0)}{dx_0^2} + C_2\frac{dq(x_0)}{dx_0} = 0. \quad (53)$$

With the boundary conditions (i) $q(0) = 1$ and (ii) $q(M) = 0$, this gives the solution

$$q(x_0) = \frac{e^{-C_2 x_0}\sinh[C_2(M - x_0)]}{\sinh(C_2 M)}. \quad (54)$$

The desired pdf is obtained as the derivative of $q(x_0)$ with respect to M,

$$P(M) = \frac{C_2 e^{-C_2 x_0}\sinh(C_2 x_0)}{\sinh^2(C_2 M)}, \quad M \geqslant x_0. \quad (55)$$

This result has been obtained in Refs. [26] and [42] in a different context. Figure 4 compares this form with numerical results at temperatures below T_m ($C_2 < 0$). We observe a good agreement for initial bubble sizes satisfying $x_0 > x_{\rm ch}$. Figure 5 further displays numerical results, both below T_m and above it, by adopting the full bubble potential.

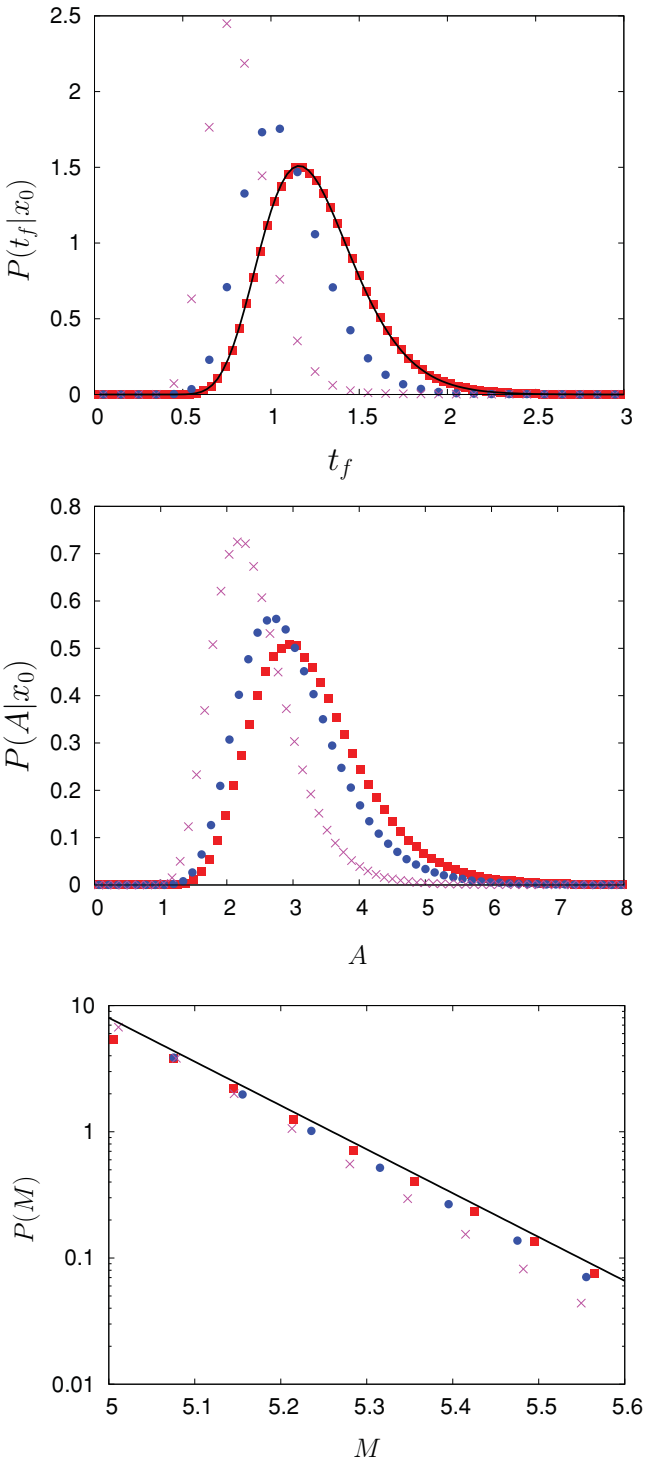

FIG. 4. (Color online) Results from numerical simulations below the melting temperature for the pdf of the first-passage time (top), the pdf of the area till the first-passage time (center), and the pdf of the maximum size until the first-passage time pdf (bottom). The parameters are $C_1 = 0$ (red squares), $C_1 = 1$ (blue dots), and $C_1 = 4$ (purple crosses). The initial bubble size is $x_0 = 5$ and $C_2 = -4$ in all the cases. The analytic results in the large bubble approximation appear in black continuous lines. In the middle panel, we do not make a comparison with the analytic function for $P(A|x_0)$ because its explicit form is known only in the limit of small and large values of A.

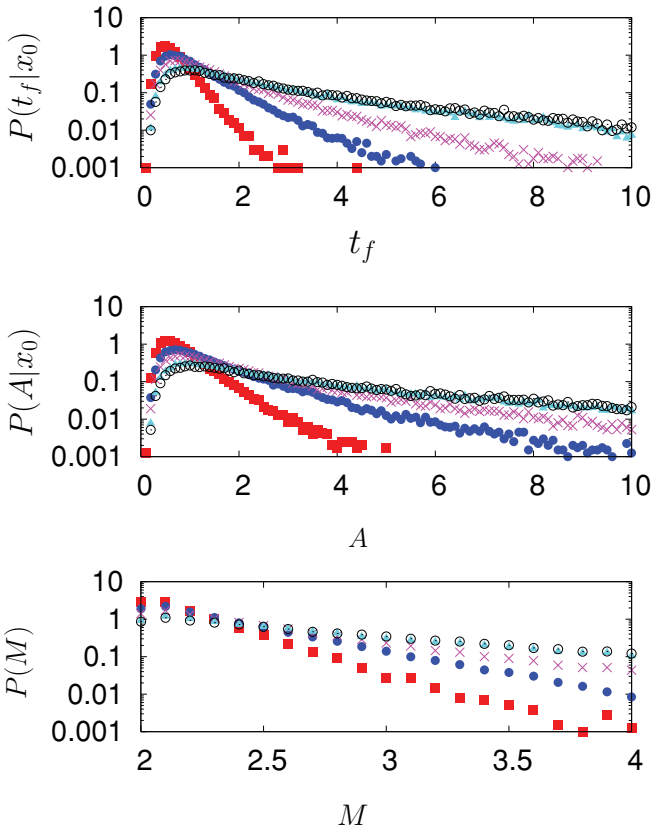

FIG. 5. (Color online) Results from numerical simulations below and above T_m, showing the first-passage time pdf (top), the pdf of the area until the first-passage time (center), and the pdf of the maximum until the first-passage time (bottom). The parameters are $C_2 = -2$ (red filled squares), $C_2 = -1$ (blue filled circles), $C_2 = -0.5$ (purple crosses), $C_2 = 0.1$ (light blue filled triangles), and $C_2 = 0.5$ (black circles). The initial bubble size is $x_0 = 2$ and $C_1 = 1$ in all cases.

D. Joint pdf of the maximum M and the corresponding time t_m before the first-passage time: $P(M,t_m)$

This distribution has been studied in [26] for the random motion (46) with $C_2 < 0$ by employing the PD method described in Sec. II D. To compute $P(M,t_m)$ for general C_2, we follow the discussion of Sec. III D and split a representative path into two parts, before and after t_m. We compute the weight of each part separately. The weight W_R of the path after t_m is given by $W_R = q(M-\epsilon)$. Using Eq. (54), and then taking ϵ to be infinitesimally small, we get

$$W_R = \frac{e^{-C_2(M-\epsilon)}\sinh(C_2\epsilon)}{\sinh(C_2M)} = \epsilon\frac{C_2e^{-C_2M}}{\sinh(C_2M)} + O(\epsilon^2). \quad (56)$$

The weight W_L of the part before t_m is obtained from Eq. (15) by substituting $C_1 = 0$,

$$W_L \propto e^{C_2(M-\epsilon-x_0)}\sum_{p=1}^{\infty} e^{-E_pt_m}\psi_p(M-\epsilon)\psi_p(x_0). \quad (57)$$

Here, $\psi_p(x)$ and $E_p(x)$ are the eigenfunctions and energies of the Hamiltonian (13) with the potential,

$$V(x) = \begin{cases} \frac{C_2^2}{2} & \text{if } 0 < x < M, \\ \infty & \text{if } x = 0, \quad x = M. \end{cases} \quad (58)$$

The normalized eigenfunctions are easily obtained as $\psi_p(x) = \sqrt{2/M}\sin(p\pi x/M)$ with the corresponding energies $E_p = C_2^2/2 + p^2\pi^2/(2M^2)$. Substituting this in Eq. (57), we get

$$W_L \propto \epsilon e^{C_2(M-x_0)-C_2^2t_m/2}\frac{2\pi}{M^2}\sum_{p=1}^{\infty}(-1)^{p+1}p\sin\left(\frac{p\pi x_0}{M}\right) \times e^{-p^2\pi^2t_m/(2M^2)} + O(\epsilon^2). \quad (59)$$

The last equation, together with Eq. (56), gives the total probability,

$$P(M,t_m;\epsilon) = B(\epsilon)W_LW_R. \quad (60)$$

The normalization constant $B(\epsilon)$ is determined by requiring that $\lim_{\epsilon\to 0}\int_0^{\infty} P(M,t_m;\epsilon)dt_m \to P(M)$, given in Eq. (55). With $\sum_{p=1}^{\infty}(-1)^{p-1}p\sin(px)/(p^2+a^2) = (\pi/2)\sinh(ax)/\sinh(a\pi)$ for $-\pi < a < \pi$ [37], one gets $B(\epsilon) = 1/2\epsilon^2$. Using this expression for $B(\epsilon)$ in Eq. (60) and taking the limit $\epsilon \to 0$, we get the desired pdf,

$$P(M,t_m) = \frac{C_2e^{-C_2x_0-C_2^2t_m/2}}{\sinh(C_2M)}\frac{\pi}{M^2}\sum_{p=1}^{\infty}(-1)^{p+1}p \times \sin\left(\frac{p\pi x_0}{M}\right)e^{-p^2\pi^2t_m/(2M^2)}. \quad (61)$$

The marginal distribution $P(t_m)$ can be obtained by integrating $P(M,t_m)$ over M from x_0 to infinity. The large-t_m and small-t_m asymptotic behaviors of $P(t_m)$ for $C_2 < 0$ are discussed in Ref. [26].

V. CONCLUSIONS

In this paper, we derived probability distribution functions of various Brownian functionals associated with a random-walk model for DNA bubble dynamics at temperatures below, at, and above the denaturation temperature. Based on the BFP method discussed in Ref. [25], we derived (i) the first-passage time distribution $P(t_f|x_0)$, providing information about the bubble lifetime, (ii) the distribution $P(A|x_0)$, of the area A covered by the random walk until the first-passage time, measuring the bubble reactivity to processes within the DNA, and (iii) the distribution $P(M)$, of the maximum bubble size M before bubble closure, all conditioned on an initial bubble of size x_0 ($x_0 \in [0,M]$). (iv) The joint probability distribution $P(M,t_m)$ of the maximum bubble size M and the time t_m of its occurrence before the first passage time was also obtained by employing the Feynman-Kac path integral formulation. The advantage of the elegant methods adopted here is that they produce results on various functionals by making proper choices of a single term in a parent differential equation with appropriate boundary conditions.

We considered separately the dynamics of small and large bubbles. Analytical results for the pdfs at each limit agree well with Langevin simulations. Our analysis reveals

TABLE I. Scaling behavior of the probability distribution functions of various Brownian functionals calculated in this work for small and large DNA bubbles.

Quantities	Results for small bubble	Results for large bubble
$P(t_f\|x_0)$	$\sim t_f^{-C_1-3/2} e^{-x_0^2/2t_f}$ $\sim t_f^{-C_1-3/2}; t_f \to \infty$ $\sim e^{-x_0^2/2t_f}; t_f \to 0$	$\sim t_f^{-3/2} e^{-(x_0+C_2 t_f)^2/2t_f}$ $\sim t_f^{-3/2} e^{-C_2 x_0} e^{-C_2^2 t_f/2}; t_f \to \infty$ $\sim e^{-(x_0+C_2 t_f)^2/2t_f}; t_f \to 0$
$C(x_0,t)$	$\sim t^{-C_1-1/2}; t \to \infty$ $\approx 1 - \frac{(x_0^2/2)^{C_1-1/2}}{\Gamma(C_1+1/2)} t^{1/2-C_1} e^{-x_0^2/2t}; t \to 0$	$\sim t^{-3/2} \exp\left\{-\left[\frac{(x_0-\|C_2\|t)^2}{2t}\right]\right\}; t \to \infty$
$P(A\|x_0)$	$\sim A^{-\frac{2}{3}(C_1+2)} e^{-2x_0^3/9A}$ $\sim A^{-\frac{2}{3}(C_1+2)}; A \to \infty$ $\sim e^{-2x_0^3/9A}; A \to 0$	$A^{-3/4} e^{-\left(\frac{8}{3}\right)^{1/2} \|C_2\|^{3/2} A^{1/2}}; A \to \infty$ $\sim A^{-4/3} e^{-2x_0^3/9A}; A \to 0$
$P(M)$	$\sim M^{-2C_1-2}$	$\sim \frac{1}{\sinh^2(C_2 M)}$
$P(M,t_m)$	$\sim M^{-(C+3)} \sum_{p=1}^{\infty} u_{\alpha p} \frac{e^{-u_{\alpha p}^2 t_m/2M^2}}{J_{\alpha+1}(u_{\alpha p})} J_\alpha\left(\frac{u_{\alpha p} x_0}{M}\right)$, where $C = 2C_1 + 1 - \sqrt{1+4C_1^2}/2$	$\sim \frac{e^{-C_2^2 t_m/2}}{M^2 \sinh(C_2 M)} \sum_{p=1}^{\infty} (-1)^{p+1} p \sin\left(\frac{p\pi x_0}{M}\right) e^{-p^2\pi^2 t_m/2M^2}$

different nontrivial scaling behaviors of $P(t_f|x_0)$, $P(A|x_0)$, $P(M)$ and $P(M,t_m)$, as summarized in Table I. The scaling exponents are characterized either by the entropic parameter C_1, or by the base-pair dissociation parameter C_2. These quantities may thus be estimated experimentally by using fluorescence correlation spectroscopy [9] to measure, e.g., the maximum size distribution $P(M)$ for small and large bubbles separately.

We expect our results to be useful in quantifying chemical processes within DNA, for example, protein binding to single-stranded DNA, and for developing a deeper understanding of polymer dynamics. It is of interest to extend our study and consider loop-loop interactions [30], and the effects of disorder and heterogeneity in predicting the kinetics of specific processes within DNA bubbles [45–48].

ACKNOWLEDGMENTS

D.S. and M.B. acknowledge support from the Connaught fund and from NSERC. S.G. thanks Amir Bar, Raphaël Chetrite, Ori Hirschberg, Satya N. Majumdar, and David Mukamel for fruitful discussions and suggestions, and gratefully acknowledges the Israel Science Foundation (ISF) for supporting his research at the Weizmann Institute of Science.

[1] J. D. Watson and F. H. C. Crick, Cold Spring Harb. Symp. Quant. Biol. **18**, 123 (1953).
[2] A. Kornberg and T. A. Baker, *DNA Replication Second Edition* (University Science Books, Sausalito, CA, 2005).
[3] M. D. Frank-Kamenetskii, Phys. Rep. **288**, 13 (1997).
[4] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[5] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Bio-Polymers* (Academic, New York, 1970).
[6] C. Danilowicz, Y. Kafri, R. S. Conroy, V. W. Coljee, J. Weeks, and M. Prentiss, Phys. Rev. Lett. **93**, 078101 (2004).
[7] N. K. Voulgarakis, A. Redondo, A. R. Bishop, and K. Ø. Rasmussen, Phys. Rev. Lett. **96**, 248101 (2006).
[8] O. Krichevsky and G. Bonnet, Rep. Prog. Phys. **65**, 251 (2002).
[9] G. Altan-Bonnet, A. Libchaber, and O. Krichevsky, Phys. Rev. Lett. **90**, 138101 (2003).
[10] B. S. Alexandrov, V. Gelev, S. W. Yoo, A. R. Bishop, K. Ø. Rasmussen, and A. Usheva, PLoS Comput. Biol. **5**, e1000313 (2009).
[11] B. S. Alexandrov, V. Gelev, S. W. Yoo, L. B. Alexandrov, Y. Fukuyo, A. R. Bishop, K. O. Rasmussen, and A. Usheva, Nucleic Acids Res. **38**, 1790 (2010).
[12] J. Phys.: Condens. Matter, **21** (2009), "Special Section on DNA melting".
[13] R. Metzler, T. Ambjörnsson, A. Hanke, and H. C. Fogedby, J. Phys. Condens. Matter **21**, 034111 (2009); and references therein.
[14] H. Kunz, R. Livi, and A. Süto, J. Stat. Mech.: Theory Exp. (2007) P06004.
[15] T. Ambjörnsson, S. K. Banik, M. A. Lomholt, and R. Metzler, Phys. Rev. E **75**, 021908 (2007).
[16] S. K. Banik, T. Ambjörnsson, and R. Metzler, Europhys. Lett. **71**, 852 (2005).
[17] A. Hanke and R. Metzler, J. Phys. A **36**, L473 (2003).
[18] A. Bar, Y. Kafri, and D. Mukamel, Phys. Rev. Lett. **98**, 038103 (2007).
[19] A. Bar, Y. Kafri, and D. Mukamel, J. Phys. Condens. Matter **21**, 034110 (2009).
[20] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
[21] S. Srivastava and Y. Singh, Europhys. Lett. **85**, 38001 (2009).
[22] H. C. Fogedby and R. Metzler, Phys. Rev. Lett. **98**, 070601 (2007); Phys. Rev. E **76**, 061915 (2007).
[23] L.-A. Wu, S. S. Wu, and D. Segal, Phys. Rev. E **79**, 061901 (2009).

[24] Note that we refer to the functionals here as "Brownian functionals," although the dynamics involves both the random delta-correlated force and a deterministic component corresponding to the bubble potential [see Eq. (1)].
[25] S. N. Majumdar, Curr. Sci. **89**, 2076 (2005).
[26] J. Randon-Furling and S. N. Majumdar, J. Stat. Mech.: Theory Exp. (2007) P10008.
[27] M. Kac, Trans. Am. Math. Soc. **65**, 1 (1949).
[28] S. N. Majumdar and M. J. Kearney, Phys. Rev. E **76**, 031130 (2007).
[29] P. L. Krapivsky, S. N. Majumdar, and A. Rosso, J. Phys. A **43**, 315001 (2010).
[30] Y. Kafri, D. Mukamel, and L. Politi, Physica A **306**, 39 (2002).
[31] E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958).
[32] J. P. Bouchaud and M. Mezard, J. Phys. A **30**, 7997 (1997).
[33] M. L. Deng and W. Q. Zhu, Phys. Rev. E **77**, 021918 (2008).
[34] As discussed in Ref. [26], one is required to impose an infinitesimal cutoff ϵ, or else an infinite number of crossings of the level M (immediately after the first contact at time t_m) would make it impossible to ensure that the path stays below M at times subsequent to t_m.
[35] A. C. Brańka and D. M. Heyes, Phys. Rev. E **58**, 2611 (1998).
[36] G. Schehr and P. Le Doussal, J. Stat. Mech.: Theory Exp. (2010) P01009.
[37] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series, and Products* (Academic, New York, 1965).
[38] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists Fourth Edition* (Academic, London, 1995).
[39] *Tables of Integral Transforms, Volume I, Based, in part, on notes left by Harry Bateman and compiled by the staff of the Bateman Manuscript Project*, edited by A. Erdélyi, M. F. Oberhettinger, and F. G. Tricomi (McGraw-Hill, New York, 1954).
[40] A. J. Bray, Phys. Rev. E **62**, 103 (2000).
[41] F. Bowman, *Introduction to Bessel Functions* (Dover, New York, 1958).
[42] M. J. Kearney and S. N. Majumdar, J. Phys. A **38**, 4097 (2005).
[43] A. P. Prudnikov, Yu. A. Brychkov, and O. I. Marichev, *Integrals and Series: Volume 2, Special Functions* (Gordon and Breach, Amsterdam, 1998).
[44] M. J. Kearney, S. N. Majumdar, and R. J. Martin, J. Phys. A **40**, F863 (2007).
[45] B. Coluzzi and E. Yeramian, Eur. Phys. J. B **56**, 349 (2007).
[46] T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, Phys. Rev. Lett. **97**, 128105 (2006).
[47] T. Ambjörnsson, S. K. Banik, O. Krichevsky, and R. Metzler, Biophys. J. **92**, 2674 (2007).
[48] M. Zoli, Phys. Rev. E **81**, 051910 (2010).