Work distribution for unzipping processes

Peter Werner and Alexander K. Hartmann
Institut für Physik, Universität Oldenburg, 26111 Oldenburg, Germany

Satya N. Majumdar
Laboratoire de Physique Théorique et Modèles Statistique, Université Paris-Saclay, 91405 Orsay CEDEX, France

(Received 17 January 2024; accepted 8 July 2024; published 7 August 2024)

A simple zipper model is introduced, representing in a simplified way, e.g., the folded DNA double helix or hairpin structures in RNA. The double stranded hairpin is connected to a heat bath at temperature T and subject to an external force f, which couples to the free length L of the unzipped sequence. The leftmost zipped position can be seen as the position of a random walker in a special external field. Increasing the force leads to a zipping-unzipping first-order phase transition at a critical force $f_c(T)$ in the thermodynamic limit of a very large chain. We compute analytically, as a function of temperature T and force f, the full distribution $P(L)$ of free lengths in the thermodynamic limit and show that it is qualitatively very different for $f < f_c$, $f = f_c$, and $f > f_c$. Next we consider quasistatic work processes where the force is incremented according to a linear protocol. Having obtained $P(L)$ already allows us to derive an analytical expression for the work distribution $P(W)$ in the zipped phase $f < f_c$ for a long chain. We compute the large-deviation tails of the work distribution explicitly. This distribution can be interpreted as work distribution for an oscillatorylike model. Our analytical result for the work distribution is compared over a large range of the support down to probabilities as small as 10^{-200} with numerical simulations performed by applying sophisticated large-deviation algorithms.

DOI: 10.1103/PhysRevE.110.024115

I. INTRODUCTION

The physical work W plays an important role for many equilibrium and nonequilibrium processes at all scales. The work W is a random variable that fluctuates from one realization of the underlying process to another. For large thermodynamic systems, typically the distribution of work converges to a δ function peaked at its average, and it is thus sufficient to compute just the average work. For smaller systems, however, the fluctuations of W around its average are also highly relevant, as illustrated by the field of stochastic thermodynamics [1,2]. Therefore, to describe systems with few degrees of freedom comprehensively, one needs to know the full distribution $P(W)$ of the work and not just its first moment. For example, the knowledge of $P(W)$ in a nonequilibrium process connecting two equilibrium states allows one to extract the free-energy difference between these two states by using the Jarzynski equality [3] or the theorems of Crooks [4].

In experiments [5–7] the processes are repeated many times, the work is measured for each execution, and a histogram of W is obtained. This allows one to approximate $P(W)$ in the high-probability region, i.e., for typical values. But one does not have access to a broader range of the support of $P(W)$ down to its tails. If the system contains more degrees of freedom, the estimation of free-energy differences without knowing the tails of $P(W)$ yields wrong results because the tails dominate the computed averages of exponentials of work. The same is true when numerical simulations of model processes are performed in a straightforward way and repeated many times in the so-called *simple sampling* scheme. Still, by using *large-deviation algorithms*, the work distribution $P(W)$ has been obtained in few cases for a larger range of its support, down to extreme tails with probabilities as small as 10^{-100}. In the most interesting case of complex interacting many-particle systems, $P(W)$ was obtained for the Ising model subject to a changing magnetic field [8] and for stretching of an RNA hairpin structure [9].

By using analytical studies, the distributions $P(W)$ over their full range of support have been obtained for some systems with few degrees of freedom. In very simple cases Gaussian distributions are obtained [10,11]. Furthermore, a particle in a cylinder with a moving piston [12] was considered. Other examples are given by two level systems [13] and single particles with a dynamics described by a Fokker-Planck equation [14,15] or, equivalently, by a Langevin equation [16–24], in one-dimensional potentials of varying shapes. In addition, the distributions of work performed by changing parameters of quantum harmonic oscillators [25–31] could be obtained analytically. Also, the cases of cyclic driving [32], of coupling to multiple reservoirs [33], quantum nonlinear oscillators [34], or quantum oscillators used as detectors [35] were considered. Furthermore, single-particle systems with stochastic driving were investigated [36,37]. Recently, such approaches were extended to single-particle quantum systems [38]. In a related context, models for single-particle heat engines have also been studied [39–42].

Here we want to go beyond the single-particle case and address the analytical calculation of the work distribution $P(W)$ for a more complex system of interacting particles, but

with a finite number N of degrees of freedom. Some works have been done in that direction. For example, in a system of several point particles coupled by harmonic springs, e.g., the Rouse polymer model, the work distribution can be obtained from the result of the single-particle Gaussian case [11]. In addition to the harmonic coupling, also the extreme case of very stiff nonlinear coupling can be solved [43]. The treatment of polymers can be extended to networks of harmonic oscillators [44]. Furthermore, for a relaxing elastic manifold [45] and for a chain of active particles coupled by harmonic springs [46], the heat statistics have been obtained analytically.

Our present work is motivated by the recent numerical study on stretching of RNA secondary structures, where $P(W)$ was obtained for hairpins. By applying large-deviation algorithms, almost the full range of support could be obtained [9]. In the present study we will investigate a simplified *zipper* model that can be used to describe either the opening and closing of RNA hairpin structures, or that of DNA double helices, under an external force. Our zipper model is a very simple one with nonharmonic interactions (and thus goes beyond Gaussian integrals in models with harmonic interactions) and yet allows for an exact analytic solution in the folded or "zipped" phase, in particular, exhibiting a non-Gaussian distribution of work $P(W)$.

In an early model for DNA unzipping in a solvent [47], the term coupling the free part of the helix to a solvent can be interpreted as the interaction of the free part with an external force f. The model allowed for an exact calculation of the partition function and an unzipping transition was observed. During the decades other variants of unzipping models [48–50] were studied, like the unzipping of DNA by pulling [51], and the denaturation of DNA under applied torque [52]. Also, a model exhibiting a first-order unzipping transition was analyzed [53–55]. Furthermore, unzipping was studied via the calculation of Lee-Yang zeros [56], which allowed the authors to obtain the large-deviation tail of the energy distribution. In some cases heterogeneous sequences were considered [57], in particular, a mapping to the disordered polymer in random media was provided [58], allowing for an analytical replica calculation. The actual dynamics of unzipping was modeled via a simple kinetic two-step process [59]; for reviews see, e.g., Refs. [60,61]. For the equilibrium models, unzipping transitions with model-dependent critical forces f_c were found and the behaviors in the zipped and the unzipped phases were described by average quantities (and sometimes also by the corresponding fluctuations). One important observable is the free length L, i.e., the unzipped part of the sequence which is next to the beginning of the sequence where the external force is applied. Note that by considering just L as an effective variable, integrating out other degrees of freedom, the model exhibits similarities to a random walker in an external field, see end of Sec. III.

In the present paper, we will use a simple version of the previous models but go beyond the calculation of averages and study the full distribution $P(L)$ of the free length in the zipped phase, in the unzipped phase, and also at the critical force f_c. This in turn will allow us to calculate analytically some moments and finally, the full distribution $P(W)$ of work for equilibrium processes where the external force is changed from 0 to $f_{\max}$ quasistatically in the region $f < f_c$ (zipped phase). We compare the analytical results to large-deviation simulations of the unzipping process and find very good agreement over about 200 decades in probability. Thus, the model constitutes a nice example where the work distribution is available for a process in a system of many interacting particles. This can be used as a starting point for similar consideration of the corresponding not-quasistatic process or other models of complex interacting particles. Also, the shape or generalizations of the obtained distribution can be used to fit to numerical data for other systems.

The paper is organized as follows. In Sec. II we present our zipper model. Then in Sec. III we solve the model exactly and show the existence of a first-order unzipping transition in the temperature-force plane and obtain, in particular, the full distribution $P(L)$ of the free length. In Sec. IV we present an exact numerical algorithm to sample configurations in equilibrium. In Sec. V we use this algorithm to confirm the analytical results for the thermodynamic behavior. Next, in Sec. VI we analytically derive, in the zipped phase, the first two moments and also the full distribution of work $P(W)$ for quasistatic processes involving a finite number of increments of the force f. In Sec. VII we evaluate these quantities in the limit of infinitely large increments. Then, in Sec. VIII we present the numerical results of the large-deviation simulations showing the work distributions over hundreds of decades in probability and compare with the analytical findings. Finally, we conclude in Sec. IX with a summary and outlook.

II. A SIMPLE UNZIPPING MODEL

We consider a zipper consisting of two complementary halves of a sequence consisting of N opposite pairs of bases where each pair can independently be bonded (close) or unbonded, see the top part of Fig. 1. Note that we assume the simplest case where each base can only be bonded to the complementary base in the pair, not to other bases.

The part of the sequence that is outside the "outmost" bonded pair, i.e., the upmost pair in the top part of Fig. 1, is denoted as *free*. We consider the case where the first base is fixed and the last base is coupled to, e.g., an optical tweezer, such that a force f can be exerted on the zipper. Thus, it is the free part of the zipper which couples to the force.

This situation can be described by a one-dimensional lattice of N sites $i = 1, \ldots, N$, which represents the first half or strand of the zipper. At each site we have a binary variable $\sigma_i \in \{0, 1\}$, indicating whether the base i is bonded ($\sigma_i = 1$) or not, respectively. A typical configuration of this binary string is shown in Fig. 1. Let M denote the total number of 1's in a configuration. We also denote by L the number of lattice sites to the left of the first 1 (appearing in the string as one reads from left to right)—this is one-half of the free length of the zipper. Evidently all these L sites contain 0 by definition, see bottom of Fig. 1. Thus a typical configuration is labeled by two integers M and L. Note that for a given L in $0 \leqslant L \leqslant N$, the number of 1's, i.e., the variable M, can take values only in the range $0 \leqslant M \leqslant N - L$. We define the energy of the configuration as [62]

$$E(M, L) = -J M - 2 f L, \tag{1}$$

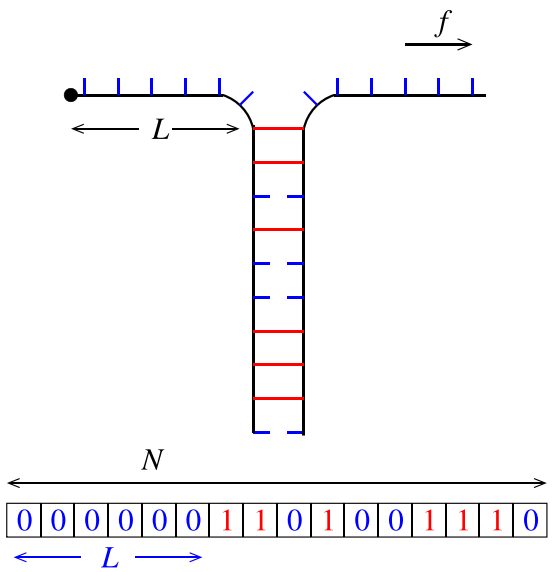

FIG. 1. Top: A sequence which is partially unzipped under an applied force f. The length of the outside unzipped part is L. Also in the zipped part, some pairs are not bonded, forming so-called bubbles. Bottom: binary-string representation of length N of this configuration consisting of 1's and 0's. Thus, L denotes the length of the substring made of consecutive 0's from the left before the first 1.

where $f > 0$ is the applied force and $J > 0$ is the binding energy of a base pair between a base in one strand and its partner base located at the equivalent position in the other strand.

Note that this model, basically equivalent to Ref. [47], is a simplified model of real RNA or DNA sequences. Actually, much more realistic models exist which, e.g., take entropic contributions into account. The most realistic models exist in the form of numerical packages like *RNA structure* [63] or the *Vienna RNA Package* [64]. In particular, when comparing to actual unzipping experiments on RNA or DNA, such packages are required. We consider it as natural to start with the most simple model, which allows us to obtain the distribution analytically over the full range of support. Somehow more sophisticated models might be considered analytically in the future, while the most realistic models are likely to be studied only by numerical approaches.

III. EQUILIBRIUM BEHAVIOR

Ground state. Let us first investigate the ground state by minimizing the energy function in Eq. (1). We need to maximize $-E(M, L) = J\,M + 2\,f\,L$. Since both terms in $-E$ are non-negative, we can maximize them one after the other. First fix L and vary M. The maximum value of M for fixed L is clearly $(N - L)$. Hence $-E(M = N - L, L) = J\,(N - L) + 2\,f\,L = J\,N + (2f - J)\,L$. We now have to maximize this function with respect to L where $0 \leqslant L \leqslant N$. There are two possibilities:

(i) The case $0 < f < J/2$: In this case the function $J\,N + (2f - J)\,L$ is maximized when $L = 0$, i.e., the first entry (from the left) in the ground-state configuration must be a 1. This is then the "zipped" phase.

(ii) The case $f > J/2$: In this case the function $J\,N + (2f - J)\,L$ is maximized when $L = N$. This means the ground state consists of all 0's. This is thus a totally "unzipped" phase.

In summary, there is a phase transition in the ground state from the "zipped" phase to the "unzipped" phase, as one increases f (fixed J) through the critical value $f_c = J/2$. We will see below that this "unzipping" phase transition persists at finite temperature also.

Finite temperature. At finite temperature we associate a Boltzmann weight $e^{-\beta\,E(M,L)}$ to each configuration labeled by (M, L), where the energy $E(M, L)$ is given in Eq. (1) and $\beta = 1/T$ is the inverse temperature. Henceforth, we fix J and consider the behavior of the system as a function of two control parameters (T, f) in the force-temperature plane. Our goal is to obtain the phase diagram in the $(T - f)$ plane. The partition function of the model is defined as

$$Z_N(\beta, f) = \sum_{\text{all config.}} e^{-\beta\,E(M,L)}. \tag{2}$$

To evaluate the partition function, we carry out the sum over configurations in two steps. We first fix L and sum over all values of M. After this, we sum over all values of L. Thus we write

$$Z_N(\beta, f) = \sum_{L=0}^{N} W_N(L), \tag{3}$$

where we define

$$W_N(L) = \sum_{M=0}^{N-L} e^{-\beta\,E(M,L)}. \tag{4}$$

When carrying out the sum over M, we need to distinguish two cases, namely, when $0 \leqslant L \leqslant N - 1$ and when $L = N$. In the latter case $(L = N)$, the string consists entirely of 0's, hence $M = 0$. In this case the Boltzmann weight factor is just $e^{2\beta f N}$. For $0 \leqslant L \leqslant N - 1$, we note that once we fix L, the $(L + 1)$-th entry is necessarily a 1. Hence, the remaining $(M - 1)$ 1's can be placed in the available $(N - L - 1)$ boxes such that each box can contain at most one 1. The number of ways this can be done is simply $\binom{N-L-1}{M-1}$. Thus, beyond the $(L + 1)$-th entry, $M - 1$ follows a Bernoulli distribution, and the net partition sum can be obtained by summing over all possible values of M. For fixed $0 \leqslant L \leqslant N$, it can be expressed as

$$W_N(L) = e^{2\beta f N}\,\delta_{L,N} + \left[\sum_{M=1}^{N-L} \binom{N-L-1}{M-1} e^{\beta J M + 2\beta f L}\right] \mathbf{1}_{0\leqslant L\leqslant N-1}, \tag{5}$$

where $\mathbf{1}_{0\leqslant L\leqslant N-1}$ is an indicator function, which is 1 if $0 \leqslant L \leqslant N - 1$, and $\delta_{L,N}$ is the Kronecker δ function. The sum over M in Eq. (5) can be performed trivially using binomial expansion, and one gets, after slight rearrangement,

$$W_N(L) = \frac{(1 + e^{\beta J})^N}{(1 + e^{-\beta J})}\,[\mu^L\,\mathbf{1}_{0\leqslant L\leqslant N-1} + (1 + e^{-\beta J})\,\mu^N\,\delta_{L,N}], \tag{6}$$

where we introduced the important parameter

$$\mu = \frac{e^{2\beta f}}{1 + e^{\beta J}}. \tag{7}$$

Thus, we have arrived at an effective one-dimensional model with L as degree of freedom and controlled by the parameter μ. Finally, summing Eq. (6) over L (it is just a simple geometric series), we get the exact partition function valid for arbitrary positive N:

$$\begin{aligned} Z_N(\beta, f) &= \sum_{L=0}^{N} W_N(L) \\ &= \frac{(1 + e^{\beta J})^N}{(1 + e^{-\beta J})} \left[\frac{1 - \mu^N}{1 - \mu} + (1 + e^{-\beta J})\, \mu^N \right], \end{aligned} \tag{8}$$

where μ is given in Eq. (7).

We now analyze the thermodynamic limit $N \to \infty$. The free energy per site is defined as

$$\mathcal{F}(\beta, f) = -\lim_{N\to\infty} \frac{1}{\beta} \frac{\ln Z_N(\beta, f)}{N}. \tag{9}$$

Taking logarithm of Eq. (8) and the $N \to \infty$ limit, we see that the limiting value of the free energy per site depends on whether $\mu \leqslant 1$ or $\mu > 1$. We obtain

$$\mathcal{F}(\beta, f) = \begin{cases} -\frac{1}{\beta} \ln(1 + e^{\beta J}) & \text{if} \quad \mu \leqslant 1 \\ -2\, f & \text{if} \quad \mu > 1. \end{cases} \tag{10}$$

Note that $\mathcal{F}$ does not depend on the force for $\mu < 1$ and is continuous at $\mu = 1$, but its first derivative with respect to T or f is discontinuous at the critical point $\mu = 1$, indicating a first-order phase transition.

To shed more light on the two phases and the transition between them, let us now define the average fraction of 1's in the string as an order parameter:

$$\langle m \rangle = \lim_{N\to\infty} \frac{\langle M \rangle}{N} = -\frac{\partial \mathcal{F}(\beta, f)}{\partial J}. \tag{11}$$

Using Eq. (10) for the free energy per site, one gets

$$\langle m \rangle = \begin{cases} \frac{e^{\beta J}}{1+e^{\beta J}} & \text{if} \quad \mu \leqslant 1 \\ 0 & \text{if} \quad \mu > 1. \end{cases} \tag{12}$$

Thus the phase $\mu > 1$ corresponds to the "unzipped" phase where $\langle m \rangle = 0$, while the phase $\mu \leqslant 1$ corresponds to the "zipped" phase where $\langle m \rangle$ is nonzero. As one approaches the critical point $\mu \to 1$ from below, the order parameter $\langle m \rangle$ undergoes a finite jump, thus confirming the first-order phase transition,

Thus the critical line in the $(T - f)$ plane is obtained by setting $\mu = 1$. Using the expression of μ in Eq. (8), one then obtains the critical curve $f_c(T)$ in the $(T - f)$ plane:

$$\mu = 1 \Longrightarrow f_c(T) = \frac{\ln(1 + e^{\beta J})}{2\beta} = \frac{T}{2} \ln(1 + e^{J/T}). \tag{13}$$

The phase diagram in the $(T - f)$ plane including the critical line $f_c(T)$ is shown in Fig. 2. Note that the critical force *increases* when increasing the temperature, i.e., the fluctuations do not help. This is known as *cold unzipping* in the literature.

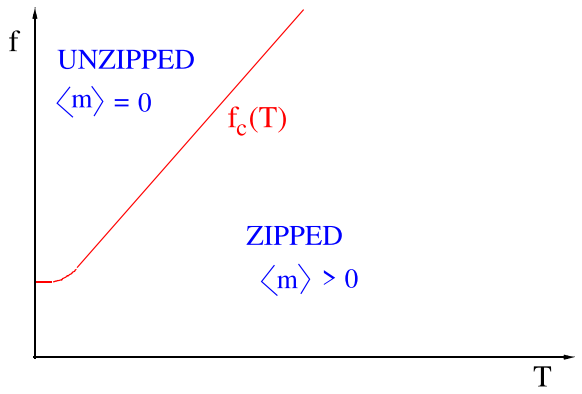

FIG. 2. Phase diagram in the $(T - f)$ (temperature-force) plane. The critical line $f_c(T) = (T/2)\ln(1 + e^{J/T})$ (drawn schematically in the figure) separates the unzipped phase $[f > f_c(T)]$ from the zipped phase $[f < f_c(T)]$. The order parameter $\langle m \rangle$, namely, the fraction of 1's in the string in the thermodynamic limit, is nonzero in the zipped phase and vanishes in the unzipped phase. On the critical line $f = f_c(T)$, the order parameter $\langle m \rangle$ is nonzero and jumps to 0 as one enters the unzipped phase from the zipped side, indicating a first-order phase transition.

The critical curve $f_c(T)$ has the following asymptotic behaviors:

$$f_c(T) \simeq \begin{cases} \frac{J}{2} + \frac{T}{2}\, e^{-J/T} - \frac{T}{4}\, e^{-2J/T} + \cdots & (T \to 0) \\ \frac{1}{2}(\ln 2)\, T + \frac{J}{4} + \frac{J^2}{16T} + O(T^{-2}) & (T \to \infty). \end{cases} \tag{14}$$

To obtain the $T \to 0$ limit, we rewrite the expression of $f_c(T)$ in Eq. (13) as $f_c(T) = J/2 + (T/2)\ln(1 + e^{-J/T})$ and then expand the logarithm in powers of $e^{-J/T}$. This gives the first line of Eq. (14). In the opposite $T \to \infty$ limit, we first expand $e^{J/T} = 1 + J/T + J^2/(2T^2) + \cdots$ and then expand the logarithm in Eq. (13) in powers of $1/T$, yielding the second line in Eq. (14). Thus, as $T \to 0$, the critical value $f_c(0) = J/2$ is consistent with the ground-state analysis before.

Statistics of the free length L. One can also characterize the transition in terms of the variable L, denoting the free length of the RNA chain. Here, we will not only calculate the average value but actually the full distribution $P(L|N)$ for any given N. This will allow us later on to obtain for slow processes the full distribution of work analytically.

Indeed, it follows from Eq. (6) that $P(L|N)$ is given exactly, for all $0 \leqslant L \leqslant N$, by

$$\begin{aligned} P_N(L) &= \frac{W_N(L)}{Z_N(\beta, f)} \\ &= \frac{\mu^L\, \mathbf{1}_{0\leqslant L\leqslant N-1} + (1 + e^{-\beta J})\, \mu^N\, \delta_{L,N}}{\frac{1-\mu^N}{1-\mu} + (1 + e^{-\beta J})\, \mu^N}. \end{aligned} \tag{15}$$

It is easy to check that $P_N(L)$ is normalized, i.e., $\sum_{L=0}^{N} P_N(L) = 1$. A plot of this distribution for a small value of N and $f > f_c$ highlighting the δ peak, is given in Fig. 3.

Interestingly, the distribution exhibits mainly an exponential behavior $\sim \mu^L$. Thus, if μ was equal to $e^{-\epsilon/T}$, with constant ϵ, then $P(L)$ would be similar to the energy distribution $\sim e^{-\epsilon L/T}$ of a quantum harmonic oscillator. For the

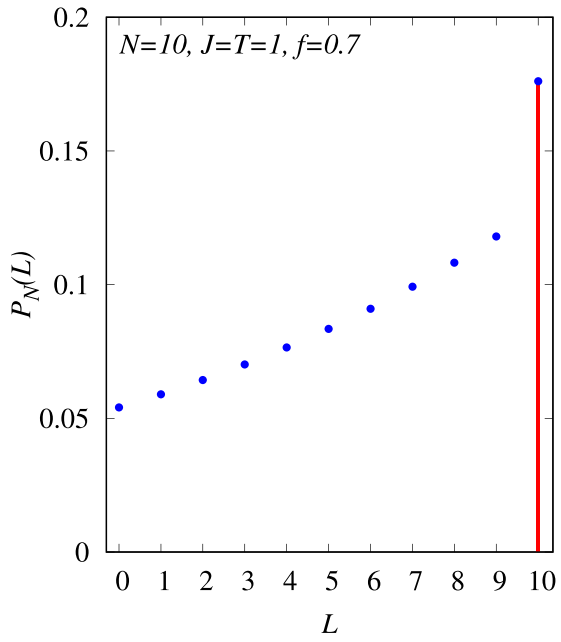

FIG. 3. Distribution of the free length L in the unzipped phase [see Eq. (15)] for $T = J = 1$, $N = 10$, and force value $f = 0.7 > f_c$, where $f_c \approx 0.6566$. The red vertical line highlights the δ peak at $L = N$.

present case, μ exhibits a much more complex dependence on the external force f and the temperature as given by Eq. (7), corresponding to changing the rigidity of a quantum harmonic oscillator. This leads in particular to strong changes of the distribution's shape around the unzipping transition, as detailed below. Such a behavior does not exist for a quantum harmonic oscillator, because at the transition the rigidity would be zero. Also, from the viewpoint of an oscillator, the roles of low- and high-energy states would be exchanged at the transition. Note that this resemblance of the distributions cannot be seen easily, considering the definition Eq. (1) of the unzipping model, and is rather a result of our calculation.

From Eq. (15), it is easy to compute its first moment for all N,

$$\langle L\rangle = \sum_{L=0}^{N} L\, P_N(L) = \frac{\frac{\mu}{1-\mu}\left[\frac{1-\mu^N}{1-\mu} - N\,\mu^{N-1}\right] + (1+e^{-\beta J})\,N\,\mu^N}{\frac{1-\mu^N}{1-\mu} + (1+e^{-\beta J})\,\mu^N}, \tag{16}$$

where μ is given in Eq. (7). In Sec. V samples for $\langle L\rangle$ as function of the force f are shown and compared to results from numerical exact sampling.

It is interesting to compute $\langle L\rangle$ in the large-N limit. It follows from Eq. (16) that this thermodynamic limit depends crucially on whether $\mu > 1$, $\mu < 1$, or $\mu = 1$. Taking this limit carefully, we find that as $N \to \infty$,

$$\langle L\rangle \simeq \begin{cases} \frac{\mu}{1-\mu} & \text{for} \quad \mu < 1 \\ \frac{N}{2} & \text{for} \quad \mu = 1 \\ N & \text{for} \quad \mu > 1. \end{cases} \tag{17}$$

Hence, as $N \to \infty$, the average free length available per site $\langle l\rangle = \langle L\rangle/N$ approaches to 0 for $\mu < 1$ (zipped phase) and 1 for $\mu > 1$ (unzipped phase). Exactly, on the critical line $\mu = 1$, we find $\langle l\rangle \to 1/2$. This statistic of $\langle l\rangle$ also confirms the first-order nature phase transition at $\mu = 1$ for a different measurable quantity.

Let us also analyze the asymptotic form of the full distribution $P_N(L)$ in Eq. (15) in the thermodynamic limit $N \to \infty$. The behavior again depends on whether $\mu < 1$ (zipped phase), $\mu > 1$ (unzipped phase), or on the critical lime $\mu = 1$ (critical).

(i) Zipped phase ($\mu < 1$). In this case, as $N \to \infty$ in Eq. (15) the distribution $P_N(L)$ converges to an N-independent form which is purely geometric. For $L = 0, 1, 2, \ldots$ one obtains

$$P_{N\to\infty}(L) = (1-\mu)\,\mu^L. \tag{18}$$

In this $N \to \infty$ limit, the average value $\langle L\rangle$ approaches a constant

$$\langle L\rangle = (1-\mu)\sum_{L=0}^{\infty} L\,\mu^L = \frac{\mu}{1-\mu}, \tag{19}$$

in accordance with the first line of Eq. (17). The fluctuations of L around this mean are quantified by the variance $\mathrm{Var}(L) = \langle L^2\rangle - \langle L\rangle^2$ that also approaches to a constant $\mathrm{Var}(L) = \mu/(1-\mu)^2$ in the large-N limit.

(ii) Critical line ($\mu = 1$). On the critical line $\mu = 1$, we find that $P_N(L)$ in Eq. (15) approaches a scaling form as $N \to \infty$,

$$P_{N\to\infty}(L) \to \frac{1}{N}\,F\left(\frac{L}{N}\right), \tag{20}$$

where the scaling function $F(x) = 1$ for $0 \leqslant x \leqslant 1$ and $F(x) = 0$ for $x > 1$. In other words, the distribution of L approaches a flat uniform distribution over $L \in [0, N]$. Consequently, the average value $\langle L\rangle$ approaches the value $N/2$ as found in the second line of Eq. (17). One can also check that the variance $\mathrm{Var}(L) = \langle L^2\rangle - \langle L\rangle^2 \to N^2/12$ as $N \to \infty$, indicating that the fluctuations of L remain big in the large-N limit at the critical point, unlike in the zipped phase where it is of $O(1)$ as $N \to \infty$.

(iii) Unzipped phase ($\mu > 1$). In this phase, the distribution $P_N(L)$ in Eq. (15) does not approach a limiting form as $N \to \infty$. Instead, there is a decaying N-dependent "bulk" part that coexists with a δ peak (condensate) at the right edge of the support at $L = N$. More precisely, for large N we find from Eq. (15) that

$$P_{N\to\infty}(L) \to \frac{(\mu-1)}{[1+(\mu-1)(1+e^{-\beta J})]} \times [\mu^{L-N}\,\mathbf{1}_{0\leqslant L\leqslant N-1} + (1+e^{-\beta J})\,\delta_{L,N}]. \tag{21}$$

Note that the weight of the δ peak or the condensate at $L = N$ approaches an N-independent value as $N \to \infty$. However, the "bulk" part of the distribution, even though it decays exponentially as $\sim\mu^{L-N}$ away

from the condensate, does actually contribute to all moments of L even in the $N \to \infty$ limit. For example, the average value $\langle L \rangle \to N + O(1)$ (as $N \to \infty$) as in the third line of Eq. (17). The variance $\mathrm{Var}(L) = \langle L^2 \rangle - \langle L \rangle^2$, however, approaches to a constant as $N \to \infty$:

$$\mathrm{Var}(L) \to \frac{\mu[1 + (\mu^2 - 1)(1 + e^{-\beta J})]}{(\mu - 1)^2[1 + (\mu - 1)(1 + e^{-\beta J})]^2}. \tag{22}$$

For both moments, the δ peak as well as the "bulk" part contributes in the large-N limit. So, one cannot neglect this "bulk" part in the thermodynamic limit. This is thus a rather unusual and interesting distribution.

Note that this large-N distribution of L can be interpreted as the distribution of the position of a biased random walker in a system of N sites, if one identifies L as the effective position of a random walker with reflecting boundary condition at $L = 0$ and $(\mu - 1)$ as the effective bias to the right. When $\mu < 1$, the drift is to the left and the position distribution of the walker reaches a stationary state as $N \to \infty$. The critical case $\mu = 1$ is like the unbiased random walker case where the position distribution becomes uniform at long time. Finally, for $\mu > 1$ there is no stationary state since the position keeps increasing with the system size N. The limiting distribution is compared for different values of f to those obtained from numerically exact sampling in Sec. V. The numerical approach is described next.

IV. SAMPLING OF CONFIGURATIONS

To sample configurations numerically, we calculate for a given system length N two conditioned partition functions $Z_f(i)$ and $Z_z(i)$, $i = 1, \ldots, N$. Here $Z_f(i)$ is the partition function of the subsequence $i \ldots N$ conditioned to the case that all sites $1, 2 \ldots i - 1$ are not bonded, i.e., on the free part. $Z_z(i)$ is the corresponding partition function for subsequence $i \ldots N$ for the zipped case, i.e., for at least one site $j \in \{1, \ldots, i - 1\}$ we have $\sigma_i = 1$. The simplest case is for $i = N$, i.e., the sequence is only the last pair. If the preceding subsequence $\sigma_1, \ldots \sigma_{N-1}$ is free, having the pair open will contribute according to Eq. (1) the energy $-2f$, while closing the pair will contribute energy $-J$. Correspondingly, if the preceding subsequence is not fully free, there will be no coupling to the force in the final site $i = N$. Therefore, having the pair open contributes energy 0, while closing the pair again yields energy $-J$. Thus, one obtains

$$\begin{aligned} Z_f(N) &= e^{2f/T} + e^{J/T} \\ Z_z(N) &= e^0 + e^{J/T}. \end{aligned} \tag{23}$$

This can be used as starting point for the recursive equations, which read as follows for $i \in \{0, \ldots, N - 2\}$:

$$\begin{aligned} Z_f(i) &= e^{2f/T} Z_f(i + 1) + e^{J/T} Z_z(i + 1), \\ Z_z(i) &= (e^0 + e^{J/T}) Z_z(i + 1) = (e^0 + e^{J/T})^{N-i}. \end{aligned} \tag{24}$$

Thus, the computation is done by the *dynamic programming* approach [65], by starting at $i = N$ and iterating site i until $i = 1$ is reached. Thus, the calculation takes $O(N)$ steps. The case $i = 1$ describes the full sequence and therefore the complete partition function is given by $Z_f(1)$, while $Z_z(1)$ is not used. Note that here it is rather easy to introduce site randomness by generalizing $J \to J_i$.

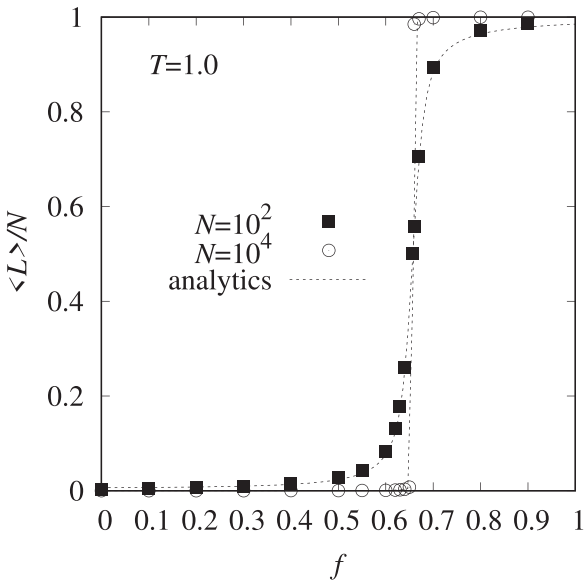

FIG. 4. Normalized expectation value $\langle L \rangle / N$ of the length of the unzipped part as a function of the force strength f, for temperature $T = J = 1$ and two different system sizes $N = 10^2, 10^4$. The symbols denote the numerical results, while the lines the analytical results from Eq. (16).

To sample a configuration $\sigma_1, \ldots, \sigma_N$ one starts at site $i = 1$, in the free part of the chain, and iteratively assigns variables σ_i, $i = 1, \ldots, N$. As long as the partial configuration is free, i.e., $\sigma_j = 0$ for all $j < i$, one assigns $\sigma_i = 0$ with probability $p_f^0(i) = e^{2f/T} Z_f(i + 1)/Z_f(i)$. Thus, with probability $1 - p_f^0(i)$ one assigns $\sigma_i = 1$. Once the first nonzero assignment $\sigma_i = 1$ has been made, one has reached the zipped part of the chain. From here one assigns $\sigma_i = 0$ with probability $p_z^0(i) = Z_z(i + 1)/Z_z(i) = 1/(1 + e^{J/T})$ and $\sigma_i = 1$ with probability $1 - p_z^0(i) = e^{J/T}/(1 + e^{J/T})$. Note that this sampling is performed in perfect equilibrium, and all sampled configurations are statistically independent for arbitrary values of temperature T and force f. To sample one configuration it takes a linear $O(N)$ number of steps. While sampling, one can directly record the size L of the free length and the number M of bonded sites.

V. NUMERICAL RESULTS FOR THE UNZIPPING TRANSITION

In Fig. 4 the result for the average length $\langle L \rangle$ is shown for $T = J = 1$ as a function of the force strength f. To investigate the finite-size effects, two different system sizes $N = 10^2$ and $N = 10^4$ are displayed. For the numerical result, an average over 10^6 randomly sampled configurations was taken for $N = 10^2$, while 10^4 configurations were considered for $N = 10^4$. An excellent agreement with the analytical results from Eq. (16) is observed. In particular, the first-order nature of the transition is very well visible for $N = 10^4$.

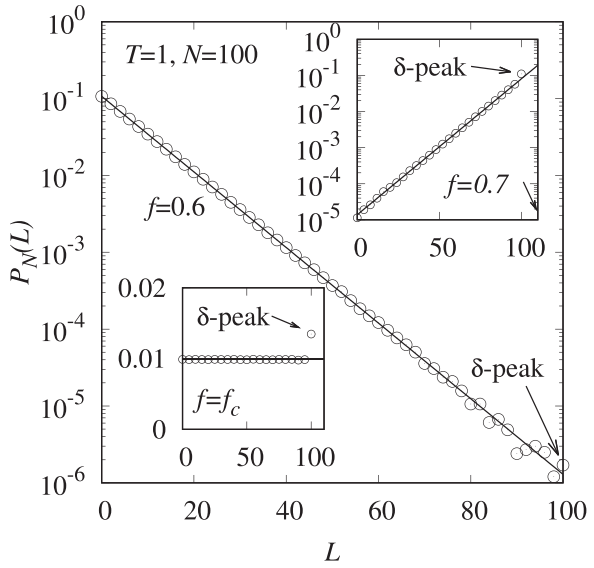

FIG. 5. Distribution of the free length L for $T = J = 1$, $N = 100$, and for force values $f = 0.6 < f_c$, $f = f_c \approx 0.6566$, and $f = 0.7 > f_c$, respectively. The symbols denote numerical results, while the lines show the analytical asymptotic distributions from Eqs. (18), (20), and (21). The δ peak at $L = N$ is for some cases only slightly visible, in particular, for $f = 0.6$, where the relative fluctuations are large and therefore highlighted in the figure by arrows, respectively.

In Fig. 5 examples for the distributions $P_N(L)$ of the free length are shown for the case $T = J = 1$ and $N = 100$. Three cases $f < f_c$, $f = f_c$, and $f > f_c$ are presented. The numerical results are histograms obtained from 10^7 independently sampled configurations. For $f < f_c$ a clear decreasing exponential function is visible. Despite the rather small system size $N = 100$, a very good match with the limiting $N \to \infty$ analytical result from Eq. (18) is visible, apart from the statistical fluctuations for large values of L. For $f = f_c$ an almost full uniform distribution is found, as obtained in Eq. (20), plus a small peak at $L = 100$, which is present in the full distribution Eq. (15), but should decrease in weight for $N \to \infty$. For $f > f_c$ a rising exponential matching the result of Eq. (21) is clearly visible. Here also a finite peak at $L = N$ appears, which should remain also for large system sizes. In general, finite-size effects appear very small. Thus, the rather small size $N = 100$ almost represents the limiting $N \to \infty$ behavior.

VI. WORK DISTRIBUTION IN THE ZIPPER MODEL IN THE ZIPPED PHASE

In this section we compute the work distribution in the zipper model in the zipped phase by increasing the applied force quasistatically. Here, we will consider increasing the applied force in discrete steps in units of $f_0 > 0$. In other words, the applied force at the mth step is given by

$$f(m) = f_0\, m \quad \text{where } m = 0, 1, 2, \ldots, N_s, \tag{25}$$

where N_s is the total number of steps. We restrict ourself in the present work to the zipped phase $\mu < 1$, because the simple shape of the distribution $P_N(L) \to (1-\mu)\,\mu^L$ (as $N \to \infty$) in this phase allows us to perform all computations analytically. Therefore, the final force $f(N_s) = f_0 N_s$ stays below the critical force $f_c = \ln(1 + e^{\beta J})/(2\beta)$ as given by Eq. (13).

To define the work appropriately, consider the following general situation. Suppose we have a system with a Hamiltonian that depends on the local degrees of freedom such as the spins $\{s_i\}$ on a lattice and also contains a parameter $\lambda(t)$ that evolves in continuous time (for example, the external magnetic field). We write this Hamiltonian as $H(\{s_i\}, \lambda(t))$. Now, we evolve the system up to a final time t_s following $\lambda(t)$. Then, quite generally, the work done to the system up to the final time t_s is defined as [3]

$$W \equiv \int_0^{t_s} dt\, \dot\lambda(t)\, \frac{\partial H[\{s_i\}, \lambda(t)]}{\partial \lambda(t)}. \tag{26}$$

For a fixed set of spins $\{s_i\}$, as λ evolves, the Hamiltonian changes, i.e., the energy changes. Hence, when integrated up to t_s as in Eq. (26), this gives the total energy pumped into or released from the system, defined as work, just due to the change of the parameter and not due to spin fluctuations.

We now adapt the general definition of work in Eq. (26) to our zipper model. Here, the external force f in the Hamiltonian in Eq. (1) plays the role of the parameter $\lambda(t)$, but we assume that our evolution occurs in discrete steps and not in continuous time. Then from Eq. (25), we get the discrete-time analog of $\dot\lambda(t)$ in Eq. (26), namely, $\dot\lambda(t) \equiv f_0$. Using, furthermore, $\partial_\lambda H \equiv -2\, L_m$, where L_m is the random variable describing the free length of the chain after the mth step, we get the discrete-time equivalent of Eq. (26):

$$W = -2\, f_0 \sum_{m=0}^{N_s-1} L_m. \tag{27}$$

We further assume that the system equilibrates after each step (quasistatic), i.e., at the mth step, the probability distribution of a configuration is given by the Boltzmann weight $\propto e^{-\beta\, E_m(M,L)}$ with inverse temperature β and with the energy function

$$E_m(M, L_m) = -J\, M - 2\, f(m)\, L_m. \tag{28}$$

It is convenient to define the rescaled work

$$w = -\frac{W}{2\, f_0} = \sum_{m=0}^{N_s-1} L_m \geqslant 0. \tag{29}$$

Our goal is to find the distribution of the random variable w in Eq. (29). To compute this, we will use the fact that L_m at the mth step is distributed via the equilibrium distribution in the zipped phase as given in Eq. (18), i.e.,

$$P(L_m) = (1-\mu_m)\, \mu_m^{L_m}, \quad \text{where } \mu_m = \frac{e^{2\,\beta\, f_0\, m}}{1 + e^{\beta J}}. \tag{30}$$

Note that we have already taken the thermodynamic limit $N \to \infty$. A consequence of Eq. (30) is

$$\langle z^{L_m} \rangle = \sum_{L_m=0}^{\infty} z^{L_m}\, P(L_m) = \frac{1-\mu_m}{1-\mu_m\, z}. \tag{31}$$

We now consider the generating function of the rescaled work w in Eq. (29) and use the fact that the L_m's for different m's

are statistically independent. This gives, using the result in Eq. (31), a rather nice and simple expression,

$$\langle z^w \rangle = \prod_{m=0}^{N_s-1} \langle z^{L_m} \rangle = \prod_{m=0}^{N_s-1} \frac{1-\mu_m}{1-\mu_m z}, \tag{32}$$

where we recall that μ_m is given in Eq. (30).

From the generating function in Eq. (32), one can easily compute all the moments and cumulants by taking derivatives with respect to z and setting $z = 1$. For this purpose, it is convenient to write $z = e^{-s}$ and derive the cumulants by taking derivatives with respect to s and set $s = 0$. Thus Eq. (32) reads, in the variable s,

$$\langle e^{-s\,w} \rangle = \prod_{m=0}^{N_s-1} \frac{(1-\mu_m)}{[1-\mu_m\, e^{-s}]}. \tag{33}$$

Consequently, the cumulant generating function is given by

$$\ln[\langle e^{-s\,w} \rangle] = \sum_{n=1}^{\infty} \kappa_n \frac{(-s)^n}{n!}, \tag{34}$$

where κ_n is the nth cumulant. Taking the logarithm of Eq. (33) and expanding for small s, one can obtain all the cumulants. For example, the first two cumulants are given by

$$\kappa_1 = \langle w \rangle = \sum_{m=0}^{N_s-1} \frac{\mu_m}{1-\mu_m}, \tag{35}$$

$$\kappa_2 = \langle w^2 \rangle - \langle w \rangle^2 = \sum_{n=0}^{N_s-1} \frac{\mu_m}{(1-\mu_m)^2}. \tag{36}$$

Let us now turn to the generating function of the full distribution in Eq. (33). First, we want to point out an interesting fact. Substituting $s = -2f_0\beta$ in Eq. (33), we get

$$\langle e^{-\beta W} \rangle = \langle e^{2\beta f_0\, w} \rangle = \prod_{m=0}^{N_s-1} \frac{(1-\mu_m)}{\left[1-\mu_m\, e^{2\beta f_0}\right]}$$
$$= \prod_{m=0}^{N_s-1} \frac{1+e^{\beta J} - e^{2\,\beta\, f_0\, m}}{1+e^{\beta J} - e^{2\,\beta\, f_0\,(m+1)}}. \tag{37}$$

However, now the cross terms in the numerator and denominator of Eq. (37) cancel telescopically, leaving behind

$$\langle e^{-\beta W} \rangle = \frac{e^{\beta J}}{1+e^{\beta J} - e^{2\,\beta\, f_0 N_s}}. \tag{38}$$

However, in the zipped phase where $\mu < 1$ it is then easy to see that in the thermodynamic limit the partition function in Eq. (8) reduces, with $f = f(m)$, to

$$Z_N(\beta,\, f(m)) \xrightarrow[N\to\infty]{} \frac{(1+e^{\beta J})^{N+1}}{(1+e^{-\beta J})[1+e^{\beta J} - e^{2\,\beta\, f_0\, m}]}. \tag{39}$$

Hence Eq. (38) can then be expressed as

$$\langle e^{-\beta W} \rangle = \frac{Z_{N\to\infty}(\beta,\, f\, N_s)}{Z_{N\to\infty}[\beta,\, f(0)]}$$
$$= \exp\left[-\beta\left(\mathcal{F}[\beta, f(m=N_s)] - \mathcal{F}[\beta, f(m=0)]\right)\right], \tag{40}$$

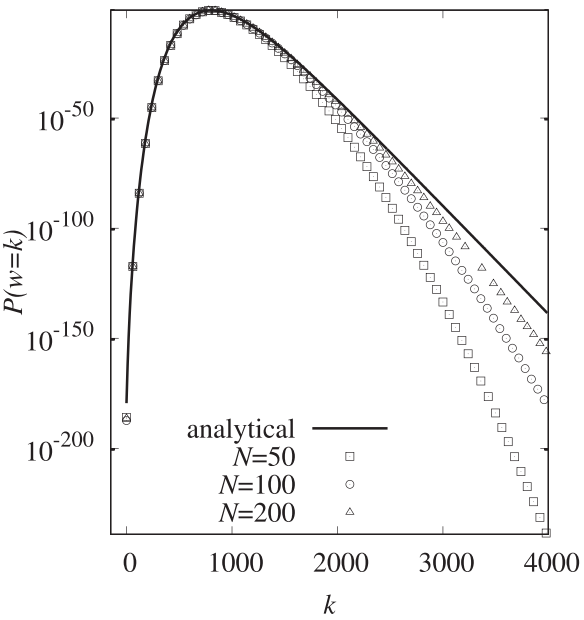

FIG. 6. Analytical $N \to \infty$ work distribution (solid line) in Eq. (43), measured in terms of integers $k = -W/(2f_0)$, for the case $\beta = J = 1$ and for $N_s = 502$. The symbols denote results obtained from numerical large-deviation sampling, as described in Sec. VIII, for three different chain lengths N. Error bars are smaller than symbol size.

where $\mathcal{F}[\beta,\, f(m)] = -(1/\beta)\ln Z[\beta,\, f(m)]$ is the free energy of the system at the mth step. In fact, Eq. (40) is nothing but the discrete version of the Jarzynski equality [3].

Note that the Jarzynski equality holds only when we set $s = -2\beta f_0$. It does not help us to compute the full work distribution. To compute this, we need to keep a general s in Eq. (33) and try to invert this generating function. In fact, since w in Eq. (29) is an integer, we first rewrite Eq. (32) as

$$\langle z^k \rangle = \sum_{k=0}^{\infty} \text{Prob.}\,[w = k|N_s]\, z^k = \prod_{m=0}^{N_s-1} \frac{(1-\mu_m)}{[1-\mu_m\, z]}. \tag{41}$$

Now, one can formally invert the generating function in Eq. (41) using Cauchy's theorem,

$$\text{Prob.}\,[w = k|N_s] = \int_C \frac{dz}{2\pi i}\, \frac{1}{z^{k+1}} \prod_{m=0}^{N_s-1} \frac{(1-\mu_m)}{[1-\mu_m\, z]}, \tag{42}$$

where C denotes a closed contour in the complex z plane around $z = 0$. The integrand in Eq. (42) has simple poles at $z_m = 1/\mu_m$. Hence, one can evaluate the contour integral by computing the residue at each pole and summing them up (with a negative sign). This gives, after some straightforward algebra, the following explicit result:

$$\text{Prob.}\,[w = k|N_s] = \left[\prod_{l=0}^{N_s-1}(1-\mu_l)\right] \sum_{m=0}^{N_s-1} \frac{\mu_m^{N_s+k}}{\prod_{n\neq m}(\mu_m - \mu_n)}, \tag{43}$$

where μ_m is given in Eq. (30). An example of the resulting distribution is shown in Fig. 6. Also, a comparison to nu-

merical large-deviation data is included, which is presented in Sec. VIII.

VII. THE LARGE-N_s BEHAVIOR OF THE MEAN, THE VARIANCE, AND THE FULL DISTRIBUTION

In this section, we will derive the behavior of the work distribution in the scaling limit when $N_s \to \infty$, $f_0 \to 0$ but with the product

$$u = 2\,\beta\, f_0\, N_s \tag{44}$$

fixed, corresponding to a constant final force. We consider below the mean, the variance, and the full distribution separately.

A. The mean

For the mean work $\langle w \rangle$, we have the exact formula in Eq. (35), namely,

$$\langle w \rangle = \sum_{m=0}^{N_s-1} \frac{\mu_m}{1-\mu_m} = \sum_{m=0}^{N_s-1} \frac{e^{2\,\beta\, f_0\, m}}{1 + e^{\beta J} - e^{2\,\beta\, f_0\, m}}, \tag{45}$$

where we recall that $e^{2\,\beta\, f_0\, N_s} < 1 + e^{\beta\, J}$. Let us define $x = 2\,\beta\, f_0\, m$. As m changes by 1, the variable x changes by $\Delta x = 2\,\beta\, f_0$. In the limit $f_0 \to 0$, this change $\Delta x \to 0$. Hence, one can replace the sum over m in Eq. (45) by an integral over x in this scaling limit. We get

$$\langle w \rangle \approx \frac{1}{2\,\beta\, f_0} \int_0^u dx\, \frac{e^x}{1 + e^{\beta\, J} - e^x} = \frac{1}{2\,\beta\, f_0} \ln\left[\frac{e^{\beta\, J}}{1 + e^{\beta J} - e^u}\right]. \tag{46}$$

Thus, the mean work, in the scaling limit where $N_s \to \infty$, $f_0 \to 0$ with $u = 2\,\beta\, f_0\, N_s$ fixed, can be expressed in a nice scaling form,

$$\langle w \rangle \approx N_s\, M_1(u = 2\,\beta\, f_0\, N_s), \tag{47}$$

where the scaling function $M_1(u)$ is given for $0 \leqslant u \leqslant u_c = \ln(1 + e^{\beta\, J})$ by

$$M_1(u) = \frac{1}{u} \ln\left[\frac{e^{\beta\, J}}{1 + e^{\beta J} - e^u}\right]. \tag{48}$$

The scaling function $M_1(u)$ is plotted in Fig. 7 and has the following asymptotic behaviors:

$$M_1(u) \approx \begin{cases} e^{-\beta\, J} & \text{as} \quad u \to 0 \\ -\frac{1}{u_c} \ln(u_c - u) & \text{as} \quad u \to u_c. \end{cases} \tag{49}$$

Thus the mean diverges very slowly (logarithmically) as $u \to u_c$ from below.

Physical significance. The scaling behavior of the mean work in Eqs. (47) and (48) has an interesting physical significance. In fact, for this model in the thermodynamics limit $N \to \infty$ the partition function at step m where $f_m = f_0\, m$ is given by Eq. (39). Let us now consider the ratio

$$\frac{Z_{N\to\infty}[\beta, f(m = N_s)]}{Z_{N\to\infty}[\beta, f(m = 0)]} = \frac{e^{\beta\, J}}{1 + e^{\beta\, J} - e^{2\,\beta\, f_0\, N_s}}, \tag{50}$$

where $f(m = 0) = 0$ and we used Eq. (39). Now consider the scaling limit $N_s \to \infty$, $f_0 \to 0$, but with the product $u = 2\,\beta\, f_0\, N_s$ kept fixed. Then the ratio in Eq. (50) reduces to

$$\frac{Z_{N\to\infty}[\beta, f(m = N_s)]}{Z_{N\to\infty}[\beta, f(m = 0)]} = \frac{e^{\beta\, J}}{1 + e^{\beta\, J} - e^u}. \tag{51}$$

Now, from Eqs. (47) and (48) we have the mean work (unscaled),

$$\langle W \rangle = -2\, f_0\, \langle w \rangle = -\frac{1}{\beta} \ln\left[\frac{e^{\beta\, J}}{1 + e^{\beta J} - e^u}\right]. \tag{52}$$

Now taking logarithm on both sides of Eq. (51), we can express the right-hand side of Eq. (52) as

$$\langle W \rangle = -\frac{1}{\beta} \ln\left[\frac{Z_{N\to\infty}[\beta, f(m = N_s)]}{Z_{N\to\infty}[\beta, f(m = 0)]}\right] = \mathcal{F}[\beta, f(m = N_s)] - \mathcal{F}[\beta, f(m = 0)]. \tag{53}$$

Thus the mean work is exactly the free-energy difference between the final equilibrium state and the initial equilibrium state, which is expected for quasistatic evolution.

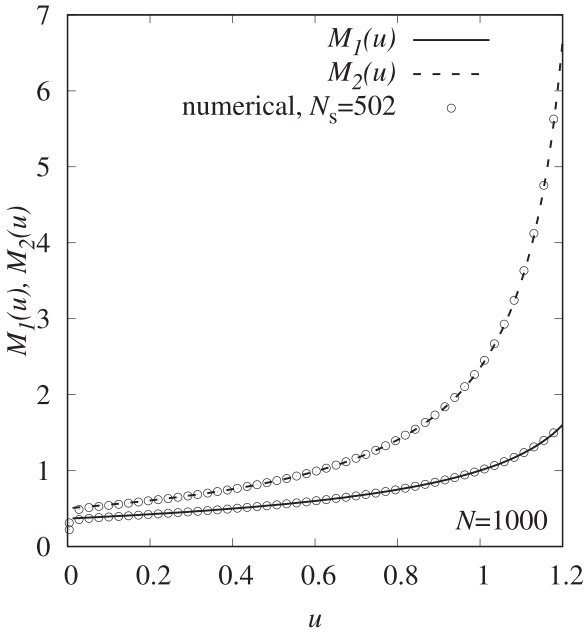

FIG. 7. The lines show the scaling functions $M_1(u)$ for the mean work of Eq. (48) and $M_2(u)$ for the variance Eq. (57). We chose $\beta = 1$, $J = 1$ and consequently, $u_c = \ln(1 + e^{\beta\, J}) \approx 1.31$. The symbols show the numerical results as presented in Sec. VIII.

B. The variance

Also for the variance, we have the exact formula in Eq. (36) that reads

$$\mathrm{Var}(w) = \langle w^2 \rangle - \langle w \rangle^2 = \sum_{m=0}^{N_s-1} \frac{\mu_m}{(1-\mu_m)^2} = (1 + e^{\beta\, J}) \sum_{m=0}^{N_s-1} \frac{e^{2\,\beta\, f_0\, m}}{\left[1 + e^{\beta\, J} - e^{2\,\beta\, f_0\, m}\right]^2}. \tag{54}$$

As in the case of the mean, we define the variable $x = 2\,\beta\, f_0\, m$, which becomes continuous in the $f_0 \to 0$ limit.

Hence, in the scaling limit in Eq. (44) we can replace the sum by an integral and write

$$\mathrm{Var}(w) \approx N_s\,(1+e^{\beta J})\,\frac{1}{u}\int_0^u dx\,\frac{e^x}{[1+e^{\beta J}-e^x]^2}. \tag{55}$$

Performing the integral explicitly, we then have the scaling behavior

$$\mathrm{Var}(w) \approx N_s\,M_2(u=2\,\beta\,f_0\,N_s), \tag{56}$$

where the scaling function $M_2(u)$ is for $0 \leqslant u \leqslant u_c = \ln(1+e^{\beta J})$, given explicitly by

$$M_2(u) = (1+e^{-\beta J})\frac{(e^u-1)}{u\,(1+e^{\beta J}-e^u)}. \tag{57}$$

The scaling function $M_2(u)$ is plotted in Fig. 7 and has the following asymptotic behaviors:

$$M_2(u) \approx \begin{cases} e^{-\beta J}(1+e^{-\beta J}) & \text{as} \quad u \to 0 \\ \frac{1}{u_c\,(u_c-u)} & \text{as} \quad u \to u_c. \end{cases} \tag{58}$$

Thus as $u \to u_c$ from below, the variance diverges as $1/(u_c - u)$, which is faster as compared to the mean in Eq. (49).

C. The full work distribution

Here, we will consider the full work distribution in the scaling limit $N_s \to \infty$, $f_0 \to 0$ with their product $u = 2\,\beta\,f_0\,N_s$ kept fixed. Our starting point is the exact Cauchy representation of the work distribution in Eq. (42), which can be rewritten as

$$\mathrm{Prob.}\,[w = k|N_s] = A\int_C \frac{dz}{2\pi i}\,\exp\left[-(k+1)\,\ln(z) - \sum_{m=0}^{N_s-1}\ln(1-\mu_m\,z)\right], \tag{59}$$

where $A = \prod_{m=0}^{N_s-1}(1-\mu_m)$, and we recall that μ_m is given by Eq. (30).

To proceed we set $k = y\,N_s$, where y is of $O(1)$ in the large-N_s limit. Taking the scaling limit as usual, i.e., by defining $x = 2\,\beta\,f_0\,m$ and replacing the sum inside the exponential by an integral, one finds, after straightforward algebra, for $y = \frac{k}{N_s}$ and $u = 2\,\beta\,f_0\,N_s$,

$$\mathrm{Prob.}\,[w = k|N_s] \approx A\;e^{N_s\,\ln(1+e^{\beta J})}\int_C \frac{dz}{2\pi i} \times \exp\left[-N_s\,S(z|y,u)\right], \tag{60}$$

where the action S is given explicitly by

$$S(z|y,u) = y\,\ln z + \frac{1}{u}\int_0^u \ln(1+e^{\beta J}-z\,e^x)\,dx. \tag{61}$$

The idea is now to perform a saddle point analysis of the integral in the large-N_s limit. Taking the derivative with respect to z and setting

$$\partial_z S(z|y,u) = 0 \tag{62}$$

gives the saddle point explicitly,

$$z^*(y|u) = \frac{(1+e^{\beta J})\,(e^{u\,y}-1)}{(e^{u\,(y+1)}-1)}. \tag{63}$$

Similarly, the prefactor $A\;e^{N_s\,\ln(1+e^{\beta J})}$ in Eq. (60) can be analyzed in the large-N_s limit, giving

$$A\;e^{N_s\,\ln(1+e^{\beta J})} \approx \exp\left[N_s\,\frac{1}{u}\int_0^u \ln(1+e^{\beta J}-e^x)\,dx\right]. \tag{64}$$

Evaluating the integral by the saddle point method for large N_s and using the result in Eq. (64), the work distribution can be expressed in a nice large-deviation form (for fixed $u = 2\,\beta\,f_0\,N_s$ and N_s large),

$$\mathrm{Prob.}\,[w = k|N_s] \approx \exp\left[-N_s\,\Phi_u\left(\frac{k}{N_s}=y\right)\right], \tag{65}$$

where the rate function $\Phi_u(y)$ is given by

$$\Phi_u(y) = y\,\ln\left[z^*(y|u)\right] + \frac{1}{u}\int_0^u \ln\left[\frac{1+e^{\beta J}-z^*(y|u)\,e^x}{1+e^{\beta J}-e^x}\right]dx, \tag{66}$$

with $z^*(y|u)$ given in Eq. (63). Performing the integral explicitly, we get

$$\Phi_u(y) = y\,\ln\left[z^*(y|u)\right] + \frac{1}{u}\left[\mathrm{Li}_2\left(\frac{z^*(y|u)}{1+e^{\beta J}}\right) - \mathrm{Li}_2\left(\frac{1}{1+e^{\beta J}}\right) + \mathrm{Li}_2\left(\frac{e^u}{1+e^{\beta J}}\right) - \mathrm{Li}_2\left(\frac{e^u\,z^*(y|u)}{1+e^{\beta J}}\right)\right], \tag{67}$$

where

$$\mathrm{Li}_2(z) = \sum_{n=1}^{\infty}\frac{z^n}{n^2} \tag{68}$$

is the dilogarithm function. A plot of the rate function $\Phi_u(y)$ vs y (for fixed u) is given in Fig. 8. By taking the derivative with respect to y (for fixed u) and setting $\partial_y\Phi_u(y) = 0$ gives the value $y_{\min}(u)$, where the rate function $\Phi_u(y)$ has its minimum. It is not difficult, then, to show that

$$y_{\min}(u) = \frac{1}{u}\,\ln\left[\frac{e^{\beta J}}{1+e^{\beta J}-e^u}\right] \equiv M_1(u), \tag{69}$$

where $M_1(u)$ defined in Eq. (47) is just $\langle w\rangle/N_s$. Indeed, one expects the rate function $\Phi_u(y)$ to be a convex function with a minimum at $y = y_{\min}(u) = M_1(u)$ and must have a quadratic form near this minimum,

$$\Phi_u(y) \approx \frac{[y-M_1(u)]^2}{2\,M_2(u)}, \tag{70}$$

where $M_2(u)$ is the scaling function associated to the variance in Eq. (56) and computed explicitly in Eq. (57). Indeed, by expanding $\Phi_u(y)$ up to quadratic order around $y = y_{\min}(u)$, one does recover $M_2(u)$ from the rate function. The asymptotic behaviors of $\Phi_u(y)$ as $y \to 0$ and $y \to \infty$ can also be deduced easily. Let us define the parameter

$$a = 1 + e^{\beta J}. \tag{71}$$

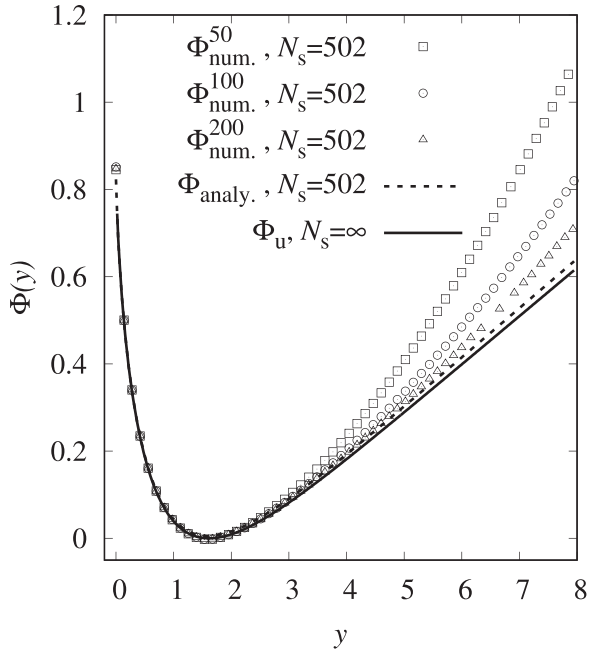

FIG. 8. Analytical (line) $N_s \to \infty$ rate function $\Phi_u(y)$ from Eq. (67) for fixed $u = 1.2$. The rate function has a minimum at $y_{\min} = M_1(u = 1.2) \approx 1.6$ and has the quadratic behavior near the minimum as in Eq. (70). Also shown is the analytical result for finite $N_s = 502$ (broken line) and numerical estimates $\Phi^N_{\text{num.}}(y)$ as obtained from large-deviation simulations, see Sec. VIII. Error bars are smaller than symbol size.

In terms of this parameter a, we can express the small- and large-y asymptotics of $\Phi_u(y)$ (for fixed u) as follows:

$$\Phi_u(y) = \begin{cases} \frac{1}{u}\left[\text{Li}_2\left(\frac{e^u}{a}\right) - \text{Li}_2\left(\frac{1}{a}\right)\right] \\ \quad + y\,\ln y + O(y), \quad \text{as } y \to 0 \\ \frac{1}{u}\left[-\text{Li}_2\left(\frac{1}{a}\right) + \text{Li}_2\left(\frac{e^u}{a}\right) + \text{Li}_2(e^{-u}) - \frac{\pi^2}{6}\right] \\ \quad + \ln(a\,e^{-u})\,y + O\left(\frac{1}{y}\right) \quad \text{as } y \to \infty. \end{cases} \tag{72}$$

Interestingly, $\Phi_u(0) = [\text{Li}_2(\frac{e^u}{a}) - \text{Li}_2(\frac{1}{a})]/u$ is a positive constant. This implies, from Eq. (65), that the probability of vanishing work, i.e., $w << N_s$, decays exponentially with increasing N_s as

$$\text{Prob.}\,[w = k|N_s] \underset{k \ll N_s}{\longrightarrow} \exp\,[-\theta(u)\,N_s],$$
$$\text{where} \quad \theta(u) = \Phi_u(0) = \frac{1}{u}\left[\text{Li}_2\left(\frac{e^u}{a}\right) - \text{Li}_2\left(\frac{1}{a}\right)\right]. \tag{73}$$

Also, since $\Phi_u(y)$ increases linearly with y for large y, see the second line of Eq. (72), it follows again from Eq. (65) that the probability of a very large work $w = k \gg N_s$ becomes independent of N_s and decays exponentially with increasing k:

$$\text{Prob.}\,[w = k|N_s] \underset{k \gg N_s}{\longrightarrow} \exp\,[-k\,\ln(a\,e^u)] = \frac{1}{(a\,e^u)^k}. \tag{74}$$

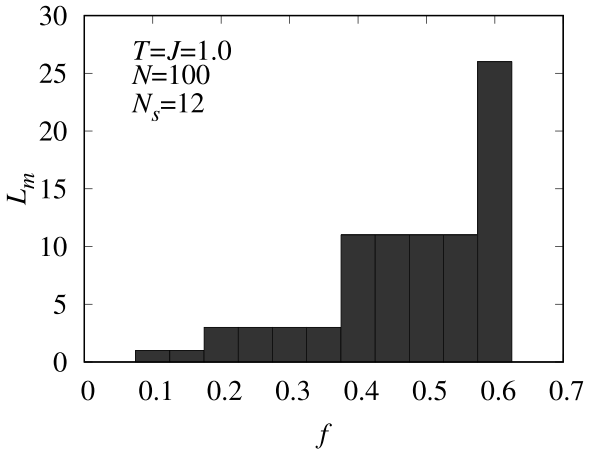

FIG. 9. A sample force-extension curve where the free length L_m at step m is shown as a function of the force f for a single random realization of the unzipping process, here for length $N = 100$, $N_s = 12$ steps, and parameters $T = J = 1$. The area under the curve is the work times a factor of -2 by Eq. (26).

VIII. LARGE-DEVIATION SIMULATIONS

Complementary to the analytical calculations of the work distribution, numerical simulations were performed, specifically to study finite-size effects. The simulation works as follows: The linear protocol is discretized into, here, $N_s = 502$ points, resulting in force values $f_m = mf_0$, $m \in [0, 1, 2, \ldots, N_s]$, where f_0 is a constant chosen such that the final force is $N_s f_0 = 0.6$, which is smaller than the critical force $f_c \approx 0.6566$ for the case $\beta = J = 1$. Therefore the phase transition should not contribute significantly. At each force value a sample from the equilibrium distribution is drawn using the algorithm as described Sec. IV, yielding a total of N_s configurations where each one exhibits a free length L_m, i.e., seen from the beginning (left), a corresponding number L_m of 0's before the first occurrence of a 1. A sample of such a *force-extension* curve is shown in Fig. 9. The work of an entire process is then given Eq. (27), which is twice the area under the curve in the figure. We considered the chain lengths $N \in \{50, 100, 200\}$.

By running the process many times, for the desired values of f_0 and N_s one can obtain histograms of the work in the high-probability region. This allowed us to measure directly the mean and the variance of the work. The resulting rescaled work and variance functions, $M_1(u)$ and $M_2(u)$, see Eqs. (47) and (56), are shown in Fig. 7. An almost perfect agreement of analytical results (for $N_s = \infty$) and numerical results is visible, except for very small values of the rescaled force u, where the finiteness of the number of steps and sequence length becomes slightly visible.

For a meaningful comparison of the full distribution to the analytical result, it is necessary to resolve the work distribution over a large range of support, down to probabilities as small as 10^{-200}. For this purpose a large-deviation algorithm [8] was employed. Note that this approach was already used to measure work distributions for unzipping processes of a more sophisticated RNA model [9]. The method's basic idea is to treat the underlying random numbers $\xi = (\xi_1, \ldots, \xi_K)$,

required by a single simulation of one full work process, as a state variable in a Markov-chain Monte Carlo process $\xi^{(1)} \to \xi^{(2)} \to \ldots$ with a target distribution that has an additional bias proportional to $\exp(W/\theta)$. Here $\theta \in [-0.0005, \ldots, 0.0005]$ is a temperaturelike parameter allowing control of the regime in which work values are predominantly generated. Since every equilibrium sample needs N random numbers and each work process requires N_s samples at different force values, the state vector ξ of random numbers has a size of $K = N \times N_s$. Each unzipping process is fully determined by ξ, and therefore the work is a deterministic function $W = W(\xi)$.

One Monte Carlo step in the Markov chain of random number vectors consists of the following steps.

(i) Randomly selecting some entries of the current state vector $\xi^{(t)}$ and redrawing them, resulting in a trial state ξ'.

(ii) Performing one full work simulation with this trial state of random numbers.

(iii) Obtaining the corresponding work $W' = W(\xi')$.

(iv) Finally accepting or rejecting the trial state with Metropolis probability $p_{\text{Metr.}} = \min\{1, \exp(\Delta W/\theta)\}$, where $\Delta W = W' - W(\xi^{(t)})$ is the work difference between the current and trial state.

Equilibration was ensured as follows: One starts the Markov chain with random as well as with extreme vectors of random numbers, such that the corresponding initial work values are very different, respectively. If for these very different initial conditions after a while the work values agree within fluctuations, the chain can be considered as equilibrated. For more details see Ref. [66]. We considered 90 different values of θ. For each one, a separate MC simulation is performed with at most 3.2×10^8 MC steps. This has yielded up to 40 000 sample points each after the initial equilibration phase, and correlations were removed from the chain of work values. Hence, for each value of θ a distribution $P_\theta(W)$ is obtained. These are then combined to the final overall work distribution via the Ferrenberg-Swendsen method [67] using a convenient tool [68].

Figure 6 shows the exact analytical and numerically estimated work distributions for the rescaled work $w = k = -W/2f_0$. The distributions could be obtained down to probability values as small as 10^{-200}. For $k < 1250$, they show good agreement for all simulated system sizes. The finite-size effects become relevant in the regime $k > 1250$. Nevertheless, the curves tend toward the analytical distribution as the system size increases.

Due to the quasistatic nature, i.e., sampling in perfect thermal equilibrium, each work value is a sum of N_s statistical independent values L_m. But these values are not identically distributed due to the different force parameter values at each of the process steps. This explains why the conditions of the central limit theorem are not fulfilled. The resulting distributions therefore do not need to be Gaussian, which they apparently are not.

An estimate of the rate functions from the numerical determined work distribution $P^N_{\text{num.}}$ is given by

$$\Phi^N_{\text{num.}}(y) = -\frac{1}{N_s} \ln\left[P^N_{\text{num.}}(k = yN_s)\right]. \tag{75}$$

Since the analytical rate function in Eq. (67) considers the case $N_s \to \infty$, it is also worth looking at

$$\Phi_{\text{analy.}}(y) = -\frac{1}{N_s} \ln(\text{Prob.}\,[w = yN_s|N_s]), \tag{76}$$

determined from the exact work distribution but for finite N_s in Eq. (43).

The different rate functions are displayed in Fig. 8. Again, the numerical curves tend towards the analytical rate functions with increasing system size, where they are matching each other for $y < 3$. Finite-size effects become relevant for $y > 3$. The analytical rate function for finite N_s is only slightly different from the $N_s \to \infty$ limiting one towards higher values of y. This indicates that the influence of a finite number of protocol points N_s is minor compared to that of a finite system size N.

IX. SUMMARY AND DISCUSSION

In summary, we have calculated analytically and verified numerically the distribution $P(W)$ of work performed in unzipping an infinitely long closed hairpin structure under an external force, applied quasistatically in the zipped phase. This model has many interacting degrees of freedom and goes beyond past analytical works where either single-particle systems or models with simple, i.e., harmonic, interactions were considered.

As one of the main steps leading to this calculation of the work distribution, we computed exactly the full equilibrium distribution $P_N(L)$ of the free length L of an infinitely long hairpin ($N \to \infty$ limit), thus going beyond previous studies that focused mainly on the average $\langle L \rangle$ as a function of system parameters. We find that the large-N distribution $P_{N\to\infty}(L)$ drastically changes its shape at the unzipping transition point $f = f_c$, and is, in particular, very broad right at the transition.

Based on this result, we were able to compute $P(W)$ in the zipped phase analytically for all W. It also allowed us to evaluate $P(W)$ numerically over its full range of support, resulting in probabilities as small as 10^{-200} for the selected value parameters. We find generally a very good agreement between analytical and numerical approaches, except expected finite-size effects which are present in the numerical simulations. There is also a dependence on the number of steps of changes of the force, but here the influence on the results is rather limited.

With respect to the comparison to experiments, the general shape is the same, i.e., a typical value and a bell-shaped decay away from it [5,6]. But it should be stressed that our result gives the distribution even down to the low-probability tails of the distribution. On the other hand, in experiments one is restricted to, say, a few thousands of repetitions, thus the tails are not accessible. Also, properties like the mean or the variance of the experimentally measured work distribution depend a lot on the details of the studied molecules, while we study a very simple chain here. Thus, a detailed comparison of our results with experiments does not seem meaningful.

As mentioned above, the distribution $P(L)$ has an exponential shape like the distribution of states of the quantum harmonic oscillator. Still, the complex dependence of the pa-

rameter μ on the force and the temperature is very different, leading to a phase transition of the unzipping process. Such a phase transition does not occur for a harmonic oscillator. Note that many processes in nature follow an exponential distribution, like radioactive decay, but there is also no connection to a harmonic oscillator.

Correspondingly, the previously calculated work distributions for the quantum harmonic oscillator obtained at low temperature are very different from the present one. Specifically, for a setup somehow close to ours, either a two-sided exponential [25], an almost Gaussian [26,27], or a two-peak δ distribution [31] were obtained. The reason for these differences is likely that the coupling to the external force of an oscillator is different from the unzipping case. Furthermore, in these works the large deviations are not discussed.

In this paper, we have restricted our calculation of $P(W)$ only in the zipped phase, for simplicity. It would be interesting to extend our calculation so that it encompasses both phases.

Also, we have considered a very simple energy model corresponding to a homogeneous chain. It would be certainly interesting to include heterogeneous chains. This has been done for the calculation of basic measurable quantities to some extent by considering disordered chains and applying the replica trick [58]. Whether this allows for a calculation of the work distribution is not clear to us. Also, for real molecules one is usually interested in actual realizations, not in disorder averages.

Finally, our results are valid for quasistatic processes. It would also be interesting to study the work distribution for a genuinely far-from-equilibrium process. One could describe them, e.g., by considering the distribution $P(L,t)$, t being a time or step counter, and describing the dynamics by allowing for transitions between neighboring states. Still, for this model the part beyond the first pair is "shielded" from the external force, i.e., is always in equilibrium. Only the free part couples to the external force and is influenced by the rate of force change. Thus, one could expect an only small influence of the speed of the process, but this remains to be verified in future work.

ACKNOWLEDGMENTS

S.N.M. acknowledges the Alexander von Humboldt Foundation for the Gay Lussac-Humboldt prize that allowed a visit to the physics department at Oldenburg University, Germany, where this work initiated. He also acknowledges the kind hospitality of the Oldenburg University. The simulations were performed at the HPC cluster ROSA, located at the University of Oldenburg (Germany) and funded by the DFG through its Major Research Instrumentation Program (INST 184/225-1 FUGG) and the Ministry of Science and Culture (MWK) of the Lower Saxony State.

[1] U. Seifert, Rep. Prog. Phys. **75**, 126001 (2012).
[2] L. Peliti and S. Pigolotti, *Stochastic Thermodynamics: An Introduction* (Princeton University Press, Princeton, NJ, 2021).
[3] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[4] G. E. Crooks, J. Stat. Phys. **90**, 1481 (1998).
[5] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Science **296**, 1832 (2002).
[6] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, Jr., and C. Bustamante, Nature (London) **437**, 231 (2005).
[7] S. Ciliberto, Phys. Rev. X **7**, 021051 (2017).
[8] A. K. Hartmann, Phys. Rev. E **89**, 052103 (2014).
[9] P. Werner and A. K. Hartmann, Phys. Rev. E **104**, 034407 (2021).
[10] T. Speck and U. Seifert, Phys. Rev. E **70**, 066112 (2004).
[11] T. Speck and U. Seifert, Eur. Phys. J. B **43**, 521 (2005).
[12] R. C. Lua and A. Y. Grosberg, J. Phys. Chem. B **109**, 6805 (2005).
[13] H. T. Quan, S. Yang, and C. P. Sun, Phys. Rev. E **78**, 021116 (2008).
[14] T. Speck, J. Phys. A **44**, 305001 (2011).
[15] B. Saha and S. Mukherji, J. Stat. Mech. (2014) P08014.
[16] A. Engel, Phys. Rev. E **80**, 021120 (2009).
[17] D. Nickelsen and A. Engel, Eur. Phys. J. B **82**, 207 (2011).
[18] D. Nickelsen and A. Engel, Phys. Scr. **86**, 058503 (2012).
[19] J. Hoppenau and A. Engel, J. Stat. Mech. (2013) P06004.
[20] C. Kwon, J. D. Noh, and H. Park, Phys. Rev. E **88**, 062102 (2013).
[21] A. Ryabov, M. Dierl, P. Chvosta, M. Einax, and P. Maass, J. Phys. A **46**, 075002 (2013).
[22] V. Holubec, M. Dierl, M. Einax, P. Maass, P. Chvosta, and A. Ryabov, Phys. Scr. **2015**, 014024 (2015).
[23] B. Saha and S. Mukherji, Eur. Phys. J. B **88**, 146 (2015).
[24] P. Chvosta, D. Lips, V. Holubec, A. Ryabov, and P. Maass, J. Phys. A **53**, 275001 (2020).
[25] S. Deffner and E. Lutz, Phys. Rev. E **77**, 021128 (2008).
[26] P. Talkner, P. S. Burada, and P. Hänggi, Phys. Rev. E **78**, 011115 (2008).
[27] V. A. Ngo and S. Haas, Phys. Rev. E **86**, 031127 (2012).
[28] M. A. A. Talarico, P. B. Monteiro, E. C. Mattei, E. I. Duzzioni, P. H. Souto Ribeiro, and L. C. Céleri, Phys. Rev. A **94**, 042305 (2016).
[29] R. Sampaio, S. Suomela, T. Ala-Nissila, J. Anders, and T. G. Philbin, Phys. Rev. A **97**, 012131 (2018).
[30] Z. Fei and H. T. Quan, Phys. Rev. Res. **1**, 033175 (2019).
[31] A. Moradian and F. Kheirandish, Phys. Scr. **96**, 125119 (2021).
[32] I. J. Ford, D. S. Minor, and S. J. Binnie, Eur. J. Phys. **33**, 1789 (2012).
[33] T. Monnai, Phys. Rev. E **81**, 011129 (2010).
[34] S. Deffner, O. Abah, and E. Lutz, J. Chem. Phys. **375**, 200 (2010).
[35] B.-M. Xu, J. Zou, and Z.-C. Tu, Commun. Theor. Phys. **73**, 065102 (2021).
[36] G. Verley, C. V. den Broeck, and M. Esposito, New J. Phys. **16**, 095001 (2014).
[37] S. Manikandan and S. Krishnamurthy, Eur. Phys. J. B **90**, 258 (2017).
[38] K. Funo and H. T. Quan, Phys. Rev. Lett. **121**, 040602 (2018).
[39] J. Hoppenau, M. Niemann, and A. Engel, Phys. Rev. E **87**, 062127 (2013).

[40] V. Holubec, J. Stat. Mech. (2014) P05022.
[41] K. Proesmans, C. Driesen, B. Cleuren, and C. Van den Broeck, Phys. Rev. E **92**, 032105 (2015).
[42] V. Holubec and A. Ryabov, J. Phys. A: Math. Theor. **55**, 013001 (2022).
[43] M. Vucelja, K. S. Turitsyn, and M. Chertkov, Phys. Rev. E **91**, 022123 (2015).
[44] M. Damak, M. Hammami, and C.-A. Pillet, J. Stat. Phys. **180**, 263 (2020).
[45] Y.-X. Wu, J.-F. Chen, J.-H. Pei, F. Zhang, and H. T. Quan, Phys. Rev. E **107**, 064115 (2023).
[46] D. Gupta and D. A. Sivak, Phys. Rev. E **104**, 024605 (2021).
[47] J. H. Gibbs and E. A. DiMarzio, J. Chem. Phys. **30**, 271 (1959).
[48] J. Applequist and V. Damle, J. Chem. Phys. **39**, 2719 (1963).
[49] U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997).
[50] D. Marenduzzo, A. Trovato, and A. Maritan, Phys. Rev. E **64**, 031901 (2001).
[51] D. K. Lubensky and D. R. Nelson, Phys. Rev. Lett. **85**, 1572 (2000).
[52] T. Hwa, E. Marinari, K. Sneppen, and L. Tang, Proc. Natl. Acad. Sci. USA **100**, 4411 (2003).
[53] Y. Kafri, D. Mukamel, and L. Peliti, Phys. Rev. Lett. **85**, 4988 (2000).
[54] Y. Kafri, D. Mukamel, and L. Peliti, Eur. Phys. J. B **27**, 135 (2002).
[55] Y. Kafri, D. Mukamel, and L. Peliti, Physica A **306**, 39 (2002).
[56] A. Deger, K. Brandner, and C. Flindt, Phys. Rev. E **97**, 012115 (2018).
[57] C. B. Roland, K. A. Hatch, M. Prentiss, and E. I. Shakhnovich, Phys. Rev. E **79**, 051923 (2009).
[58] Y. Kafri and A. Polkovnikov, Phys. Rev. Lett. **97**, 208104 (2006).
[59] A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, and D. Branton, Phys. Rev. Lett. **90**, 238101 (2003).
[60] S. Cocco, J. F. Marko, and R. Monasson, Comp. Rend. Phys. **3**, 569 (2002).
[61] P. Rissone and F. Ritort, Life **12**, 1089 (2022).
[62] M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67 (2002).
[63] J. Reuter and D. H. Mathews, BMC Bioinf. **11**, 129 (2010).
[64] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker, Algorithms Mol. Biol. **6**, 26 (2011).
[65] A. K. Hartmann, *Big Practical Guide to Computer Simulations* (World Scientific, Singapore, 2015).
[66] A. K. Hartmann, Eur. Phys. J. B **84**, 627 (2011).
[67] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[68] P. Werner, A software tool for "gluing" distributions, arXiv:2207.08429.