Pulling Pinned Polymers and Unzipping DNA

David K. Lubensky* and David R. Nelson†
Department of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 31 March 2000)

We study a class of micromanipulation experiments, exemplified by the pulling apart of the two strands of double-stranded DNA. When the pulling force is increased to a critical value, an "unzipping" transition occurs. For random DNA sequences with short-ranged correlations, we obtain exact results for the number of monomers liberated and the specific heat, including the critical behavior at the transition. Related systems include a random heteropolymer pulled away from an adsorbing surface and a vortex line in a type II superconductor tilted away from a fragmented columnar defect.

PACS numbers: 87.15.–v, 05.10.Gg, 68.35.Rh, 87.80.Fe

Recent years have seen an explosion in the use of single molecule techniques to probe biological and other "soft" materials. It is now possible, for example, to monitor the breaking of individual "lock and key" bonds [1] and the unfolding of individual proteins [2]; the mechanical properties of single DNA molecules [3] and the behavior of single molecular motors [4] have been characterized in great detail. In contrast to more traditional experiments, these new approaches give access to fluctuations on the scale of individual molecules, without the requirement for averaging over a macroscopic sample. One can, moreover, now push or pull directly on a micron-sized object, and watch how it responds. The potentially profound implications both for complex fluids and for biological physics—where single molecule techniques can often more closely mimic conditions in the cell than conventional assays—are only beginning to be explored.

Despite a number of notable contributions, theory has often been outpaced by these rapid experimental advances. Certainly, the tools available to analyze single-molecule experiments have not yet reached the level of sophistication and generality of theories of mesoscopic quantum systems. This is especially true when it comes to the role of quenched randomness, which, though often present, is typically neglected in initial theories of a given system. This Letter seeks to fill some of this gap. We study a class of micromanipulation experiments in which a polymer or other linelike object is pulled away from a confining potential well. An example of such a situation is the pulling apart of the two strands of double-stranded DNA (dsDNA) (Fig. 1). Formally, the distance between the two strands may be viewed as the coordinate of a single polymer, and the base-pairing interactions between complementary strands as a potential well. At a critical value of the pulling force, a novel phase transition occurs in which the two strands are pulled completely apart. Aspects of this transition for a homopolymer (or, equivalently, DNA with all base pairs the same) have been studied in a related model of a flux line in a type II superconductor [5,6]. Here, we show that the transition is markedly different for random heteropolymers. In particular, the number of monomers liberated as the transition is approached diverges much more strongly for heteropolymers than for homopolymers; similar differences appear in the specific heat. We calculate *exact* critical exponents and crossover functions for the random case.

Figure 1 sketches the DNA-opening experiment: One of the two single strands of a dsDNA molecule is attached to a glass slide, while a constant force $\boldsymbol{F}$ directed away from the slide acts on the end of the other strand. Methods for exerting a constant force on the piconewton scale have been developed by several groups [3,7]. Under the influence of the force $\boldsymbol{F}$, the DNA partially "unzips" at one end, breaking m base pairs. In thermal equilibrium, the degree of opening m is, of course, a fluctuating quantity. Because the base sequence of protein-coding DNA appears to many statistical tests to be random and uncorrelated along the backbone (at least up to a length scale set by the sequence's mosaic structure) [8], the free energy landscape in which m fluctuates can be taken to have a

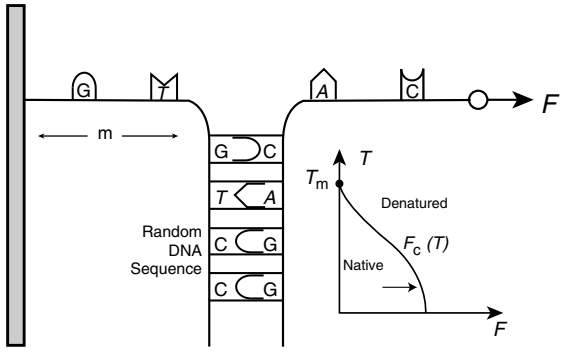

FIG. 1. Sketch of the "unzipping DNA" experiment. One of the single strands of a dsDNA molecule with a random base sequence is attached to a glass slide and the other is pulled away from the slide with a constant force F. As a result, the two strands partially separate, breaking m bonds ($m = 2$ in the figure). Inset: Schematic phase diagram in the temperature-pulling force (T–F) plane for a dsDNA molecule in three dimensions. At low T and F, the polymer is in the native, double-stranded state. At the phase transition line $F_c(T)$, the DNA denatures and the two strands separate. As indicated by the arrow, here we consider the *unzipping* transition that occurs when the transition is approached away from the $F = 0$ *melting* temperature T_m.

 0031-9007/00/85(7)/1572(4)$15.00

quenched random component. Bockelmann, Essevaz-Roulet, and Heslot have performed an elegant series of experiments in a different statistical ensemble, measuring the average force required to hold the positions of both single strands fixed [9,10]. However, because of subtleties associated with the thermodynamic limit in a single molecule system (see below), the two ensembles are not equivalent.

In the remainder of this paper, we first introduce a coarse-grained model for the interaction of the two single strands of the dsDNA. By focusing on the *unzipping* transition induced by pulling on the single strands, we can avoid treating most of the degrees of freedom explicitly, obtaining a problem that can be solved exactly by a mapping to a Markov process. Although for concreteness we will focus primarily on the DNA-opening realization of our model, our results also apply to a number of related physical systems, some of which will be described in the conclusion. Throughout, we set $k_B = 1$.

The bulk *melting* transition of dsDNA (see inset of Fig. 1) can be described by a Peyrard-Bishop–like model [11]. One views the two single strands as Gaussian polymers whose nth monomers have positions $\boldsymbol{r}_1(n)$ and $\boldsymbol{r}_2(n)$. Below the melting transition, it should be possible to neglect non-native base pairings. The interactions between the two strands, coarse grained over a number of bases, can then be described by a phenomenological potential energy term $V_n[\boldsymbol{r}(n)] = [1 + \tilde{\eta}(n)]h[\boldsymbol{r}(n)]$. Here h is a short-ranged attractive potential whose strength is temperature dependent, $\boldsymbol{r} \equiv \boldsymbol{r}_1 - \boldsymbol{r}_2$, and the variation with base sequence of the strength of the attraction between strands is described by $\tilde{\eta}(n)$, which we take to be a random variable with short-ranged correlations. The effective Hamiltonian then becomes, up to an uninteresting center of mass term,

$$\mathcal{H}_{\text{melt}} = \int_0^N dn \left\{ \frac{Td}{2b^2} \left(\frac{d\boldsymbol{r}}{dn} \right)^2 + V_n[\boldsymbol{r}(n)] \right\}, \quad (1)$$

where d is the spatial dimension and b is $\sqrt{2}$ times the Kuhn length of single-stranded DNA. Standard arguments show that the partition function $Z_{\text{melt}} \equiv \int \mathcal{D}[\boldsymbol{r}(n)] \exp(-\mathcal{H}_{\text{melt}}/T)$ obeys an (imaginary) time-dependent Schrödinger equation.

The unzipping transition may be studied by adding to $\mathcal{H}_{\text{melt}}$ the term $\mathcal{H}_{\text{pull}} = \boldsymbol{F} \cdot \boldsymbol{r}(0) = \boldsymbol{F} \cdot \boldsymbol{r}(N) - \int_0^N dn\, \boldsymbol{F} \cdot d\boldsymbol{r}/dn$. If the term proportional to $\boldsymbol{r}(N)$, which should not affect the opening near $n = 0$ for a sufficiently long polymer, is dropped, a time-dependent version of the "non-Hermitian quantum mechanics" studied in [6] results. In the time-independent case (corresponding to pulling apart homopolymeric dsDNA) there is a sharp first order unzipping transition at a critical value of the pulling force F_c satisfying $\epsilon_0(T) = -F_c^2 b^2/(2dT)$, where $\epsilon_0 < 0$ is the ground state energy of the Hermitian quantum mechanics problem obtained by setting $F = 0$. In general, $-F^2 b^2/(2dT)$ is the free energy per monomer of the unzipped monomers aligned with the pulling force. The free energy per monomer of the dsDNA that has remained zipped is ϵ_0, independent of F. The physical interpretation of the unzipping transition is thus clear: For $F < F_c(T)$, the DNA minimizes its free energy by remaining in the double-stranded form, while for $F > F_c$ it is advantageous to pull apart as many bases as possible. As $F \to F_c^-$, the free energy difference between the bound, double-stranded phase and the pulled out, single-stranded phase becomes very small, and thermal fluctuations unbind a large number of monomers near the end of the DNA. As the transition is approached, the equilibrium ensemble average of the number m of monomers that are pulled out diverges like

$$\langle m \rangle \sim (F_c - F)^{-1} \qquad \text{(homopolymer)}. \quad (2)$$

Similarly, $\langle (m - \langle m \rangle)^2 \rangle \sim (F_c - F)^{-2}$ near the transition. The thermal fluctuations about $\langle m \rangle$ are thus comparable to $\langle m \rangle$ itself. The divergence in (2) is analogous to the divergence in interface height near a wetting transition.

We now determine how results such as (2) are modified for a random DNA sequence. Sequence randomness is at worst a marginal perturbation at the ($F = 0$) melting transition in three dimensions [11–13]. The application of a Harris-like criterion [14], however, shows that the same cannot be true for the unzipping transition: The typical variation per monomer in the base-pairing energy of a pulled out section of length $\langle m \rangle$ scales like $\langle m \rangle^{-1/2} \sim \sqrt{F_c - F}$, which vanishes more slowly as the transition is approached than the relevant energy difference $|\epsilon_0| - F^2 b^2/(2dT) \sim F_c - F$. To determine the correct critical behavior, we focus on the free energy cost of pulling out a given monomer. Define $\mathcal{E}(m)$ to be the free energy of a dsDNA molecule, subject to an applied force F, of which exactly the first m monomers are unzipped. The change in $\mathcal{E}$ from pulling out one additional monomer should have the form

$$\frac{d\mathcal{E}}{dm} = f + \eta(m), \quad (3)$$

which may be integrated twice to obtain the partition function $Z = \int_0^\infty dm \exp[-\mathcal{E}(m)/T]$. Here f is the average free energy difference between an unzipped and a bound pair of complementary monomers. It vanishes like $F_c - F$ near the transition, and reduces to the familiar $|\epsilon_0| - F^2 b^2/(2dT)$ in the absence of sequence randomness. The additional term $\eta(m)$ takes account of sequence-dependent deviations from the average; it reflects the bare sequence [described by $\tilde{\eta}(m)$], dressed by thermal fluctuations. As long as $\tilde{\eta}(m)$ is a random variable with only short-ranged correlations, it is reasonable to expect that $\eta(m)$ should also be short-range correlated, with a correlation length on the order of the typical size ξ of the regions of local melting of the dsDNA strand [15]. On long enough scales, we can then take η to be Gaussian white noise, with correlator $\overline{\eta(m)\eta(m')} = \Delta\delta(m - m')$, where the overbar indicates a "disorder average" over the possible

realizations of the quenched random base sequence. The parameters f and Δ may be calculated from the $F = 0$ partition function $Z_{\rm melt}$, for example in a low temperature expansion. The model summarized by Eq. (3) can also be derived from a discrete, Ising-like description of the ds-DNA [16], and it still holds both when the single strands that have been liberated are characterized as freely jointed or wormlike chains and when there are significant excluded volume interactions [17].

The study of the unzipping transition can thus be reduced to that of a single coordinate m in the random potential $\mathcal{E}(m)$. One immediate consequence is that there is no large parameter that defines a thermodynamic limit, and thus no equivalence between the ensemble considered here and the conjugate ensemble in which m is held fixed. Below the unzipping transition, $f > 0$, and $\mathcal{E}(m)$ diverges with probability unity as $m \to \infty$. In the ensemble studied here, the unzipping fork is thus always confined to the vicinity of $m = 0$. In the absence of randomness, the probability of unzipping m monomers is $(f/T)\exp(-mf/T)$, and one recovers, e.g., (2). If there is sequence randomness, the typical random contribution to $\mathcal{E}(m)$ is of order $\sqrt{\Delta m}$; the random part thus exceeds the average contribution fm, which is responsible for the confinement, for $m \lesssim \Delta/f^2$. This length scale diverges faster than $1/f$ as $f \to 0$, suggesting that a typical value of m might show a $1/f^2$ divergence instead of the nonrandom $1/f$, with a crossover at $f \approx \Delta/T$.

Because of the confinement near $m = 0$, the unzipping transition does not exhibit self-averaging (see below). Nonetheless, one can still calculate averaged quantities and distributions over the possible realizations of randomness. To do this, one wishes to study the disorder-averaged free energy $-T\,\overline{\ln Z}$. The partition function $\tilde{Z}(m) \equiv \int_0^m dm' \exp[-\mathcal{E}(m')/T]$ of a finite-sized system of m (bound or liberated) monomers satisfies

$$\frac{d\tilde{Z}}{dm} = e^{-\mathcal{E}(m)/T} \quad \text{and} \quad \tilde{Z}(0) = 0\,; \tag{4}$$

Z follows simply by taking the limit of an infinitely long polymer: $Z \equiv \lim_{m\to\infty}\tilde{Z}(m)$. Together, Eqs. (3) and (4) form a system of coupled Langevin equations, analogous, for example, to those describing the Brownian motion of a massive particle, with $\mathcal{E}$ playing the role of momentum and $\tilde{Z}$ that of position. The associated Fokker-Planck equation for the joint distribution $P(\mathcal{E}, \tilde{Z}; m)$ of $\mathcal{E}$ and $\tilde{Z}$ at "time" m follows in the usual manner [18]:

$$\frac{\partial P}{\partial m} = \left[\frac{\Delta}{2}\frac{\partial^2}{\partial \mathcal{E}^2} - f\frac{\partial}{\partial \mathcal{E}} - e^{-\mathcal{E}/T}\frac{\partial}{\partial \tilde{Z}}\right]P\,. \tag{5}$$

By Laplace transforming with respect to $\tilde{Z}$ and to m, one can solve (5) and obtain an exact expression for the partial distribution $\int d\mathcal{E}\, P(\mathcal{E}, \tilde{Z}; m \to \infty)$ in terms of modified Bessel functions of order $2fT/\Delta$; $-T\,\overline{\ln Z}$ and other thermodynamic quantities then follow by integration.

The device of treating the quenched randomness as a Langevin noise thus leads to a number of exact results. In particular, the average degree of opening $\overline{\langle m\rangle} = -T\,\partial\,\overline{\ln Z}/\partial f$ satisfies

$$\overline{\langle m\rangle} = \frac{2T^2}{\Delta\Gamma(2fT/\Delta)}\int_0^\infty dy\, y^{2fT/\Delta - 1}(\ln y)^2 e^{-y} - \frac{2T^2\Gamma'(2fT/\Delta)^2}{\Delta\Gamma(2fT/\Delta)^2}\,, \tag{6}$$

where $\Gamma'(z) = d\Gamma(z)/dz$. The small f behavior of Eq. (6) is given by

$$\overline{\langle m\rangle} \simeq \frac{\Delta}{2f^2} \sim (F_{\rm c} - F)^{-2} \quad \text{(random heteropolymer)}\,, \tag{7}$$

confirming our expectations for a crossover from a $1/f$ to a $1/f^2$ power law when $f \sim \Delta/T$. Similarly, the singular part of the heat capacity associated with the unzipping transition, $C \sim \partial^2\,\overline{\ln Z}/\partial T^2$, crosses over from a $1/f^2$ to a $1/f^3$ divergence. One can also compute the disorder-averaged values of higher cumulants of m. For small f, $\overline{\langle m^2\rangle - \langle m\rangle^2} = T\,\partial^2\,\overline{\ln Z}/\partial f^2 \sim 1/f^3$. Unlike in the nonrandom case, the square root of this quantity diverges more slowly than $\overline{\langle m\rangle} \sim 1/f^2$, indicating that thermal fluctuations in m for a given realization of the quenched randomness (and thus for a given heteropolymer) typically become small compared to the mean value as the transition is approached.

A real space renormalization group approach to the model of Eqs. (3) and (4) due to Le Doussal, Monthus, and Fisher [19] gives further insight into the unzipping transition. This technique gives leading order results in the limit $f \to 0$, where the authors have argued that it should be exact. In this limit, it allows one to calculate the distribution $Q(\langle m\rangle)$ of thermal average values over different realizations of randomness. This takes the form of a scaling function of $\langle m\rangle f^2/\Delta$:

$$Q(\langle m\rangle) = \frac{f^2}{\pi\Delta}\, e^{-[\langle m\rangle f^2]/2\Delta}\int_0^\infty dw\, e^{-[w\langle m\rangle f^2]/2\Delta}\frac{\sqrt{w}}{w+1}\,. \tag{8}$$

This distribution yields the same asymptotic behavior of $\overline{\langle m\rangle}$ near the unzipping transition as the Fokker-Planck approach. It also predicts that $\overline{(\langle m\rangle - \overline{\langle m\rangle})^2} \sim 1/f^4$; $\langle m\rangle$ for a polymer with a given random sequence of base pairs can thus deviate significantly from the disorder average, and this system is not self-averaging.

For the randomness-dominated critical properties reported here to be observable, the variance Δ in the base-pairing energy must be sufficiently large. Then at the crossover from nonrandom to random behavior, f will also be large, and the typical value of $\langle m\rangle \sim T^2/\Delta$ will be small enough that finite-size effects do not become an issue. In this respect, dsDNA appears to be a very good candidate system. Under physiological conditions, the

difference in free energy of binding between polymers with only A-T base pairs and those with only G-C base pairs is of order T, meaning that $\langle m \rangle$ is only a few monomers when the crossover from pure to random behavior occurs.

In summary, we have described a randomness-dominated unzipping transition of dsDNA, obtaining exact expressions for the critical behavior and for the crossover from random to nonrandom scaling. Most notably, we find that when the base sequence is random and has only short-ranged correlations, the average degree of opening $\langle m \rangle$ diverges like $1/(F_c - F)^2$ as the pulling force F approaches a critical value F_c, in marked contrast to the $1/(F_c - F)$ divergence found when all of the base pairs are identical. It should be possible to arrive at analogous results for the case of DNA whose base sequence has long-ranged correlations (as may be the case for noncoding DNA [8]). If a typical variation about the average energy grows like m^β, then balancing this against mf suggests $\langle m \rangle \sim 1/f^{1/(1-\beta)}$; the short-range-correlated case is recovered when $\beta = 1/2$. The biological significance of our results remains to be determined: Processes such as DNA replication and recombination often involve unzipping of the dsDNA. Usually, however, this is accomplished by a molecular motor relying on an outside energy source, so nonequilibrium effects must be considered. More generally, the dynamics of the unzipping transition is an open question.

We have focused on the case of unzipping DNA, but our results hold equally well for a number of more conventional condensed matter systems described by the Hamiltonian $\mathcal{H}_{\rm melt} + \mathcal{H}_{\rm pull}$ [17]. The pulling of a Gaussian random *heteropolymer* away from an adsorbing surface is a natural extension of recent work on homopolymers [20]. Other examples include a heteropolymer under tension pinned to a bulk defect [21], a magnetic flux line in a type II superconductor confined to a fragmented columnar pin and subject to a transverse field [6,12], and a simplified model of the corner wetting transition in two dimensions [22]. Related models are likely to be relevant to the transverse surface magnetization and surface specific heat near the Bose glass transition of a bulk superconductor [6] and to adhesion in a random environment [23].

It is a pleasure to thank D. Branton, D. S. Fisher, and T. Hwa for helpful conversations and T. Hwa for bringing Ref. [24] to our attention. This research was supported by the NSF through Grant No. DMR97-14725 and through the Harvard MRSEC via Grant No. DMR98-09363.

Note added.—After this work was submitted for publication, we learned that related results had been obtained, in a different physical context, for a discrete version of Eqs. (3) and (4) [24].

*Email address: lubensky@cmt.harvard.edu

†Email address: nelson@cmt.harvard.edu

[1] V. Moy, E.-L. Florin, and H. E. Gaub, Science **264**, 257 (1994); E.-L. Florin, V. Moy, and H. E. Gaub, *ibid.* **264**, 415 (1994); R. Merkel *et al.,* Nature (London) **397**, 50 (1999).

[2] M. Rief *et al.,* Science **276**, 1109 (1997); M. Kellermayer *et al., ibid.* **276**, 1112 (1997); L. Tskhovrebova *et al.,* Nature (London) **387**, 308 (1997).

[3] S. B. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992); P. Cluzel *et al., ibid.* **271**, 792 (1996); S. B. Smith, Y. Cui, and C. Bustamante, *ibid.* **271**, 795 (1996); T. R. Strick *et al., ibid.* **271**, 1835 (1996).

[4] A. D. Mehta *et al.,* Science **283**, 1689 (1999), and references therein.

[5] S. M. Bhattacharjee, e-print cond-mat/9912297.

[6] N. Hatano and D. R. Nelson, Phys. Rev. Lett. **77**, 570 (1996); Phys. Rev. B **56**, 8651 (1997).

[7] D. Wirtz, Phys. Rev. Lett. **75**, 2436 (1995); F. Amblard *et al.,* Rev. Sci. Instrum. **67**, 818 (1996).

[8] C.-K. Peng *et al.,* Nature (London) **356**, 168 (1992); A. Arneodo *et al.,* Phys. Rev. Lett. **74**, 3293 (1995); S. V. Buldyrev *et al.,* Phys. Rev. E **51**, 5084 (1995).

[9] B. Essevaz-Roulet, U. Bockelmann, and F. Heslot, Proc. Natl. Acad. Sci. U.S.A. **94**, 11 935 (1997); U. Bockelmann, B. Essevaz-Roulet, and F. Heslot, Phys. Rev. Lett. **79**, 4489 (1997); Phys. Rev. E **58**, 2386 (1998).

[10] This ensemble has also been studied theoretically without sequence randomness. See J.-L. Viovy *et al.,* C. R. Acad. Sci. Paris **317**, 795 (1994); R. E. Thompson and E. D. Siggia, Europhys. Lett. **31**, 335 (1995); M. Peyrard, *ibid.* **44**, 271 (1998).

[11] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989); D. Cule and T. Hwa, *ibid.* **79**, 2375 (1997).

[12] D. Ertaş, Phys. Rev. B **59**, 188 (1999).

[13] S. M. Bhattacharjee and S. Mukherji, Phys. Rev. Lett. **70**, 49 (1993); Phys. Rev. E **48**, 3482 (1993); H. Kallabis and M. Lässig, Phys. Rev. Lett. **75**, 1578 (1995).

[14] A. B. Harris, J. Phys. C **7**, 1671 (1974).

[15] Because ξ diverges as the melting transition is approached, our description thus breaks down very near this transition, leading to effects that are beyond the scope of this Letter.

[16] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).

[17] D. K. Lubensky and D. R. Nelson (to be published).

[18] H. Risken, *The Fokker-Planck Equation* (Springer-Verlag, Berlin, 1984).

[19] P. Le Doussal, C. Monthus, and D. S. Fisher, Phys. Rev. E **59**, 4795 (1999).

[20] See, e.g., X. Châtellier *et al.,* Europhys. Lett. **41**, 303 (1998); T. J. Senden, J.-M. di Meglio, and P. Auroy, Eur. Phys. J. B **3**, 211 (1998); X. Châtellier and J.-F. Joanny, Phys. Rev. E **57**, 6923 (1998); B. J. Haupt, J. Ennis, and E. M. Sevick, Langmuir **15**, 3886 (1999).

[21] J.-U. Sommer and A. Blumen, Phys. Rev. Lett. **3**, 439 (1997).

[22] A. Lipowski, Phys. Rev. E **58**, R1 (1998).

[23] We thank S. Nagel for conversations on this point.

[24] T. Hwa and K. Sneppen (to be published); C. de Calan *et al.,* J. Phys. A **18**, 501 (1985), and references therein.