Rare-Event Induced Binding Transition of Heteropolymers

Lei-Han Tang[1] and Hugues Chaté[2]
[1]*Department of Physics, Hong Kong Baptist University, Kowloon Tong, Kowloon, Hong Kong, China*
[2]*CEA—Service de Physique de l'Etat Condensé, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette, France*

(Received 19 July 2000)

Sequence heterogeneity broadens the binding transition of a polymer onto a linear or planar substrate. This effect is analyzed in a real-space renormalization group scheme designed to capture the statistics of rare events. In the strongly disordered regime, binding initiates at an exponentially rare set of "good contacts." Renormalization of the contact potential yields a Kosterlitz–Thouless-type transition in any dimension. This and other predictions are confirmed by extensive numerical simulations of a directed polymer interacting with a columnar defect.

DOI: 10.1103/PhysRevLett.86.830 PACS numbers: 64.60.Ak, 64.70.Pf, 68.35.Rh, 87.15.Aa

The binding transition of a polymer onto another extended object, such as a second polymer, a membrane, or an interface, is of interest in a variety of physical circumstances [1]. The problem has been studied in the context of wetting in two dimensions [2,3], the depinning of a flux line from a columnar defect in type-II superconductors [4], the denaturation of double-stranded DNA molecules [5,6], and the localization of a copolymer at a two-fluid interface [7]. Faceting of vicinal surfaces can also be seen as the binding of surface steps [8].

The mathematical framework for the homopolymer binding transition is well established. For a heteropolymer (i.e., a disordered sequence of two or more letters), the transition is expected to be broadened by fluctuations in the monomer contact energy with the target object. Perturbative treatments of the disorder run into difficulties when the order of the transition becomes different from that of the pure case. In the most studied case of a directed polymer interacting with a columnar defect, *weak disorder* is found to have no effect on either the transition temperature or the critical exponents when the transverse dimension d is between one and three [2,3]. At the marginal dimension $d = 1$, Forgacs *et al.* carried out a perturbative analysis to all orders in the strength of the disorder and found only a subleading order correction to the singular part of the free energy at the transition of the pure case [2]. The perturbation theory was reexamined by Derrida *et al.* [3]. By casting the perturbative expansion in a renormalization group (RG) form, they showed that disorder shifts the transition temperature and also modifies the critical behavior, though it was not possible to predict the new critical exponents within their perturbative scheme. Attempts in settling the controversy numerically have yielded conflicting conclusions [3,6,9].

In this Letter we report a novel RG approach to the heteropolymer binding, focusing on the non-Gaussian statistics of rare attractive segments that initiate the transition. The frequency for these "good contacts," in the case of a random sequence, is described by an exponential tail in the distribution of the contact (free) energy whose decay rate q depends on temperature. When q reaches a critical value, the energy gain at these segments is sufficient to offset the logarithmic entropy cost to form a bound phase. This offers an extremely robust and generic mechanism of heteropolymer binding transition at *strong disorder* in any dimension. The critical behavior around the transition is analyzed in a real-space RG calculation. Combined with extensive simulation results on the directed random heteropolymer, we conclude that, in the disorder dominated regime, the binding transition becomes infinite order, bearing many of the characteristics of the Kosterlitz-Thouless-Berezinskii (KTB) transition in the two-dimensional *XY* model [10]. The functional RG scheme introduced here also offers fresh insight into the crossover from the weak to strong disorder regimes (controlled by other hyperbolic fixed points), to be discussed in a future publication [11].

To set the scene, let us recall the directed polymer (DP) problem on a two-dimensional square lattice (r, t) with a columnar defect at $r = 0$. For simplicity, we restrict the transverse displacement of the DP at each step $\Delta t = 1$ to three possible values, $\Delta r = -1, 0, 1$. A step on the defect at $r = 0$ picks up an energy $\eta(t)$ which depends on the location of the contact (or monomer index) t. In the case of a random heteropolymer, the contact energies $\eta(t)$ are taken to be identically distributed and independent from each other.

The above model is easily generalized to the case of a directed walk on a $(d + 1)$-dimensional hypercubic lattice, interacting with a line defect. The role of extra transverse dimensions is best appreciated in the "necklace" representation [1], where the partition sum of a DP of length L is expressed as

$$Z(L) = \sum_n \sum_{0 \le t_1 < t_2 < \cdots < t_n \le L} \exp[-H(\{t_i\})/T]. \quad (1)$$

Here t_i is the monomer index of the ith contact, and

$$H(\{t_i\}) = \sum_i \eta(t_i) + \sum_i u(t_{i+1} - t_i). \quad (2)$$

The "pair potential" $u(t)$ is obtained by summing over thermal paths in the bulk connecting two successive contacts separated by a distance t. It has the generic form

$$u(t)/T = f_0 t + \beta(d)\ln t + \mu + O(t^{-1}), \quad (3)$$

where f_0 is the reduced free energy per unit length in the bulk, $\beta(d) = 1 + \frac{1}{2}|d - 2|$ is a universal exponent characterizing the first return probability of a random walker in d dimensions, and μ is a model-specific parameter [12].

As emphasized by Fisher [1], the contact point representation (1) unifies a large class of polymer binding problems including, under certain approximations, polymers that are self-avoiding or interacting with a surface. The linear term in t in (3) drops out when the ratio $Z(L)/Z_0(L)$ is considered, where $Z_0(L) = \exp(-f_0 L)$ is the partition function of a free polymer. The dominant interaction at large distances is given by the logarithmic term. Since the density of contacts vanishes at a continuous transition, the critical properties are governed by the value of β, which serves to define universality classes. For this class of models, it suffices to consider the DP problem at $d \leq 2$, which is the case we focus on below.

Previous analytical studies, based on a perturbative treatment of the contact potential, ran into difficulties when disorder becomes relevant [2,3,9]. To elucidate how strong disorder modifies the transition, we examine first an extreme case where the contact potential is infinite everywhere except on a small number of randomly distributed sites of density A. The reduced contact potential $x_i = \eta(t_i)/T$ on these sites is assumed to be negative and exponentially distributed with a probability density function (PDF) $p(x) = q\exp(qx)$, where q is the decay rate. For $q < 1$, the one-contact partition sum $z = \sum_{1\leq i\leq L} \times \exp(-x_i)$ of a polymer of length $L \gg 1/A$ is dominated by the strongest binding site (i.e., glassy). The typical value of the potential on this site is $x_m \simeq -q^{-1}\ln(LA)$. The excess reduced free energy of a single contact is then estimated to be

$$\Delta F/T \simeq x_m + \beta\ln L = (\beta - q^{-1})\ln L + \text{const}. \quad (4)$$

Hence, for $q < q_c = 1/\beta$, a sufficiently long polymer will make use of the rare occurrence of relatively large binding energy (of order $\ln L$) on a single monomer to form the bound phase. By setting $\Delta F \simeq T$, we estimate a "localization length" $\xi_\parallel \simeq \exp(\text{const}/|q - q_c|)$. Polymers of length $L < \xi_\parallel$ are typically free, while those with $L > \xi_\parallel$ are typically bound.

Although real polymers seldom (if at all) qualify for the above description at the monomer scale, it turns out that the exponential tail can be self-generated under coarse graining. This should come across as no surprise by noting the following. (i) For a random heteropolymer, a sequence of consecutive negative η's appears spontaneously. The binding energy is proportional to the sequence length n which follows the exponential distribution. (ii) Above the transition temperature T_c, $\langle Z^q\rangle/Z_0^q$ diverge with polymer length for $q > q(T)$, while those with $q < q(T)$ remain finite. This implies that the PDF of Z has a power-law tail $Z^{-q(T)-1}$ towards large values. Equivalently, the PDF of the reduced free energy, $F/T = -\ln Z$, decays exponentially at rate $q(T)$.

The task now is to devise a suitable framework to track the renormalization of the contact energy distribution due to multiple contacts. We have succeeded in achieving this end under a real-space RG scheme. The main ideas are outlined below. The contact point representation (2) defines a one-dimensional lattice gas model. Consider a segment of b sites along the chain which is mapped to a single site under a block transformation. The site is occupied (i.e., contains a contact) if at least one of the b sites on the segment is occupied, and empty otherwise. In the former case, the Boltzmann weight of a contact, $w_i = \exp[-\eta(t_i)/T]$, transforms as

$$w = \prod_{i=1}^{b}(1 + w_i) - 1. \quad (5)$$

Rescaling $t \to bt$ reduces w by a factor b^β due to the logarithmic interaction. Hence the following iterative RG equation is seen to hold:

$$1 + \tilde{w} = \frac{1}{b^\beta}\prod_{i=1}^{b}(1 + w_i) + 1 - \frac{1}{b^\beta}. \quad (6)$$

For $b = 2$, Eq. (6) coincides with the iterative equation for the partition function $Z = 1 + w$ on the Berker lattice given by Derrida *et al.* [3]. A full treatment of the problem can be formulated in terms of the PDF $P(x)$ of the reduced contact free energy $x \equiv -\ln(1 + w)$. From the distribution of the sum $y = \sum_i x_i$,

$$Q(y) = \int dx_1 \cdots dx_b\, \delta\left(y - \sum_{i=1}^{b} x_i\right)\prod_{i=1}^{b} P(x_i), \quad (7a)$$

we obtain the distribution of $\tilde{x} \equiv -\ln(1 + \tilde{w})$,

$$\tilde{P}(\tilde{x}) = Q[\psi(\tilde{x})]\frac{d\psi}{d\tilde{x}}. \quad (7b)$$

Here $\psi(\tilde{x}) = \tilde{x} - \beta\ln b - \ln[1 + (b^{-\beta} - 1)\exp(\tilde{x})]$ is obtained from (6) by solving for y.

In the limit $b \to 1$, Eqs. (7) define a RG flow in the space of distributions. Since $Q(x)$ is a convolution of $P(x)$, it is convenient to introduce the Laplace transform,

$$\hat{P}(s) = \int_{-\infty}^{0} dx\, e^{-sx}P(x).$$

Collecting terms to the first order in $dl \equiv \ln b$, we obtain

$$\frac{d\hat{P}(s)}{dl} = \hat{P}(s)\ln\hat{P}(s) - \beta s[\hat{P}(s) - \hat{P}(s - 1)]. \quad (8)$$

Consider first the homopolymer problem described by $\hat{P}(s) = \exp(-as)$. The parameter a, which defines an effective free energy on a given scale, satisfies

$$da/dl = a + \beta(1 - e^a). \quad (9)$$

The stable fixed point at $a = 0$ corresponds to the unbound phase where the contact potential renormalizes to infinity. The binding transition occurs at $a = a_c < 0$.

The PDF $P(x) = Aq\exp(qx) + (1 - A)\delta(x)$ considered in the single-contact model above corresponds to

$$\hat{P}(s) = \begin{cases} 1 + As/(q - s), & s < q\,; \\ \infty, & s \geq q\,. \end{cases} \qquad (10)$$

Setting $A = 0$, we obtain a family of fixed-point functions of (8), parametrized by q. Writing $\hat{P}(s) = \langle Z^s \rangle$, where $Z = 1 + w$ represents a certain restricted partition function, we see that these fixed points describe correctly the behavior of the system above the binding transition.

It can be shown that, within the class of functions that have a power-law singularity at $s = q$, (10) is the only one that is renormalizable [11]. We have derived a two-parameter RG flow equation from (8) by matching terms that diverge as $s \to q$. For $A \ll 1$, the result reads

$$dq/dl = -\frac{1}{2}\,qA\,, \qquad (11a)$$

$$dA/dl = (1 - q\beta)A + \frac{2 - q}{2q}A^2. \qquad (11b)$$

The RG flow obtained by integrating (11) at $\beta = 3/2$ is illustrated in Fig. 1(a), where q is plotted against $A^{1/2}$ in analogy with the KTB transition. The critical manifold ($T = T_c$) is indicated by the thick line. For $T > T_c$, the flow ends on the $A = 0$ axis at a $q > q_c = 1/\beta$. In contrast, below but close to T_c, A first decreases, reaches a minimum value, and then increases again. Integrating the flow equation near $q = q_c$ and $A = 0$ yields an exponentially diverging correlation length $\xi_\parallel \simeq \xi_0 \exp(\alpha|T - T_c|^{-1/2})$. As in the KTB transition, the free energy per unit length has an essential singularity when approaching T_c from below,

$$f_s \simeq T/\xi_\parallel = (T/\xi_0)\exp(-\alpha|T - T_c|^{-1/2})\,. \qquad (12)$$

We have checked the above predictions against (i) full iteration of Eqs. (7) at $b = 2$ and $\beta = 3/2$ (Berker lattice with a branching ratio $n = 2\sqrt{2}$) and (ii) transfer matrix calculation in $(1 + 1)$ dimensions. In both cases, the contact energy is taken to be $\eta = u_0 + 2g^2 + 2g\epsilon$, where u_0 is the strength of the contact potential at the homopolymer binding transition and ϵ is a random variable satisfying the normal distribution. With this choice, the transition temperature of the annealed problem (as defined by $F_a \equiv -T\ln\langle Z \rangle$) is fixed at $T_a = 1$. The strength of the disorder can be tuned by varying g.

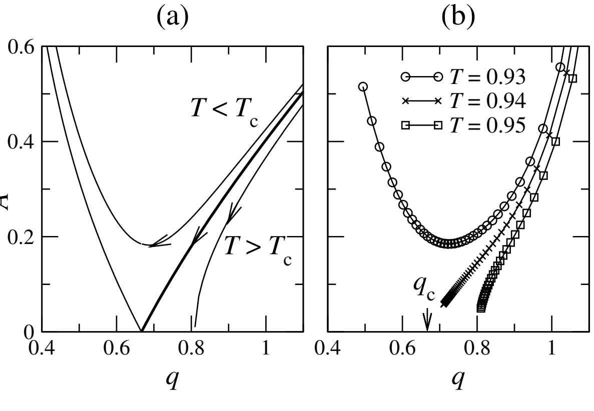

FIG. 1. RG flow from Eqs. (11) (a) and from full iteration of the distribution on the Berker lattice (b) (see text).

To probe the effective contact potential on a given scale L on the square lattice, we have computed the partition sum $Z(L)$ over all directed paths that start at $(0, 0)$ and end at any point on the $t = L$ line. The quantity $x(L) \equiv -\ln[Z(2L)/Z(L)]$, computed for a given realization of the disorder, yields a single-scale contact potential [11]. Figure 2(a) shows the integrated distribution $\Pi(x)$ for two temperatures at $g = 2$ and a sample size 10^5. The exponential tail quickly develops as L increases. At $T = T_a$ (top), the tail decays as $\exp(x)$. The preexponential amplitude decreases as a power of L. At $T = T_c$ (bottom), the decay rate of the tail approaches the critical value $q_c = 2/3$, while the amplitude changes very slowly as compared to higher temperatures. In both cases, the typical value of $x(L)$ approaches $\frac{1}{2}\ln 2$ (dotted lines) expected for a uniform repulsive potential. The integrated distribution of $x = -\ln Z$ on the Berker lattice, obtained through numerical iteration of (7), is shown in Fig. 2(b). A similar trend is observed, but the data span over a much bigger range and are much less noisy.

Using the Berker lattice data, we have fitted the tail of $\Pi(x)$ to the exponential form $A\exp(qx)$. The evolution of $A(L)$ and $q(L)$ in the critical region is shown in Fig. 1(b). Each data set corresponds to an increasing series of L in

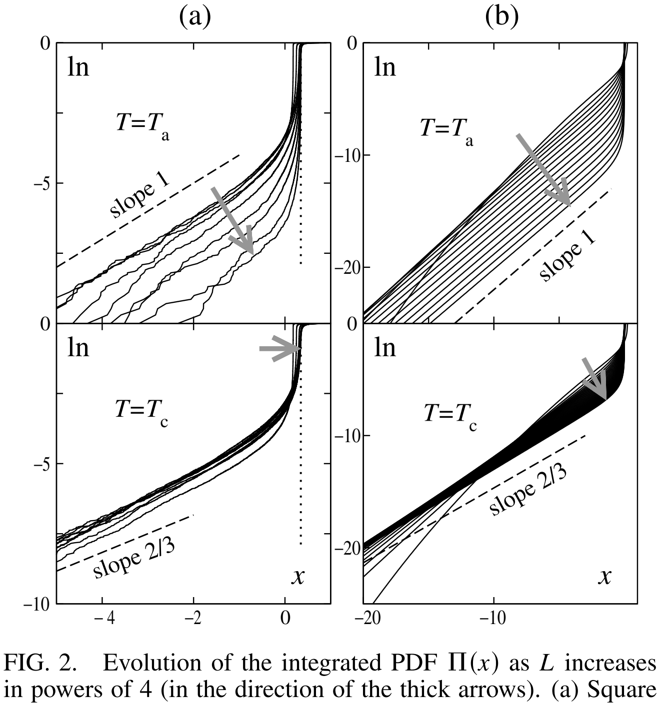

FIG. 2. Evolution of the integrated PDF $\Pi(x)$ as L increases in powers of 4 (in the direction of the thick arrows). (a) Square lattice with periodic boundary conditions and a lateral size of 2×1024 sites ($g = 2$, 10^5 samples). (b) Berker lattice at $n = 2\sqrt{2}$ and $g = 1$.

powers of 4 (from right to left) at a given temperature. The resemblance to the RG flow diagram in Fig. 1(a) is apparent. Therefore, the two-parameter flow equations (11), although only approximate in construct, correctly capture the asymptotic renormalization of the tail.

On the square lattice, the localization length $\xi_\parallel$ in the bound phase can be determined from the mean free energy $-\langle \ln Z(L) \rangle$ which goes through a maximum at $L = \xi_\parallel$, as shown in Fig. 3(a). For a polymer of length $L < \xi_\parallel$, $-\langle \ln Z \rangle$ increases with L and approaches the asymptotic behavior $\frac{1}{2} \ln L$ (dashed line) of a purely repulsive line. Thus, even at $T < T_c$, a sufficiently short polymer (or a short segment of a long polymer) is typically repelled by the target object. The attraction, due to rare events, is felt (in a typical realization of the disorder) only when $L > \xi_\parallel$. The asymptotic behavior at $L \gg \xi_\parallel$ is given by $\langle \ln Z(L) \rangle \simeq -fL$, where $f(T)$ is the reduced free energy per unit length in the thermodynamic limit. The variation of $\xi_\parallel$ and f with T, as shown in Fig. 3(b), is in agreement with Eq. (12). Extrapolating the curves to the horizontal axis (as indicated by the dashed lines), we obtain $T_c/T_a = 0.96$, 0.9, 0.85, and 0.8 for $g = 1.5$, 2, 2.5, and 3, respectively. These values agree with the analysis based on the evolution of the tails of the distribution. Details of our study will be published elsewhere.

In summary, we propose that the random heteropolymer binding transition at strong disorder is induced by rare fluctuations of the contact potential. A real-space RG scheme is presented to analyze the evolution of the PDF of the contact potential under successive block transformations. Within a two-parameter ansatz for the tail of the distribution, a RG flow similar to that of the KTB transition is derived. The analysis yields an exponentially diverging correlation length and an essential singularity for the free energy as T_c is approached from the bound side. These predictions, which hold in all dimensions when disorder is relevant, are confirmed by extensive numerical calculations on the Berker lattice and on the square lattice in (1 + 1) dimensions.

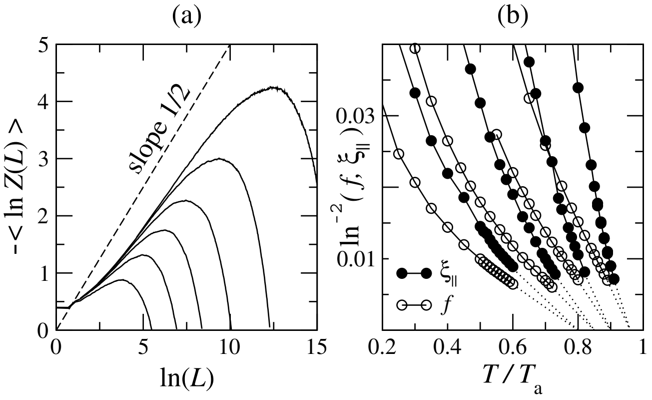

FIG. 3. Transfer matrix results for the square lattice. (a) $-\langle \ln Z \rangle$ vs $\log(L)$ for $g = 3$ and $T = 0.15$, 0.25, 0.35, 0.45, 0.55, and 0.65 (from bottom to top). (b) Variation of $\xi_\parallel$ and f with T for $g = 1.5$, 2, 2.5, and 3 (from right to left).

Finally, we comment briefly on the relationship between our work and previous numerical studies [3,6,9] of the DP binding transition on the square lattice, which were done on binary disorder with $\eta = \rho \pm \Delta$. By looking at the evolution of $P(x)$ at $T = T_a$, we discover that, even for $\Delta = \infty$ (while keeping $\rho - \Delta$ finite), the tail develops at a much slower rate than for Gaussian disorder. Since $d = 1$ is marginal with regard to the relevance of weak disorder, we expect results reported previously to be influenced, to a lesser or greater extent, by crossover effects. Indeed, the transition temperatures T_c determined in these studies, using methods different from ours, are either indistinguishable or only slightly below T_a of the corresponding annealed problem. We leave a detailed discussion of this and other issues to a future publication.

We thank the CEA Saclay (L. H. T.) and the Hong Kong Baptist University (H. C.) for hospitality, where the work was performed. The research is supported by the Hong Kong/France PROCORE program and by the Hong Kong Baptist University under Grant No. FRG/99-00/II-53.

---

[1] M. E. Fisher, J. Stat. Phys. **34**, 667 (1984).
[2] G. Forgacs, J. M. Luck, Th. M. Nieuwenhuizen, and H. Orland, Phys. Rev. Lett. **57**, 2184 (1986).
[3] B. Derrida, V. Hakim, and J. Vannimenus, J. Stat. Phys. **66**, 1189 (1992).
[4] D. R. Nelson and V. M. Vinokur, Phys. Rev. B **48**, 13 060 (1993).
[5] D. Poland and H. A. Scheraga, *Theory of Helix-Coil Transitions in Biopolymers* (Academic Press, New York, 1970).
[6] D. Cule and T. Hwa, Phys. Rev. Lett. **79**, 2375 (1997).
[7] T. Garel, D. A. Huse, L. Leibler, and H. Orland, Europhys. Lett. **8**, 9 (1989).
[8] A. Pimpinelli and J. Villain, *Physics of Crystal Growth* (Cambridge University Press, Cambridge, England, 1998).
[9] H. Kallabis and M. Lassig, Phys. Rev. Lett. **75**, 1578 (1995).
[10] J. M. Kosterlitz and D. J. Thouless, J. Phys. C **6**, 1181 (1973); J. M. Kosterlitz, J. Phys. C **7**, 1046 (1974); V. L. Berezinskii, Zh. Eksp. Teor. Fiz. **61**, 1144 (1971) [Sov. Phys. JETP **34**, 610 (1972)].
[11] L.-H. Tang and H. Chaté (to be published).
[12] The μ term in (3) can be absorbed in η. It influences the transition temperature but not the critical properties.