

Bethe approximation for a DNA-like self-avoiding walk model with variable solvent quality

D. P. Foster and C. Pinettes

Laboratoire de Physique Théorique et Modélisation, CNRS UMR 8089, Université de Cergy-Pontoise, 2 ave A. Chauvin 95302 Cergy-Pontoise Cedex, France

(Received 4 March 2009; published 8 May 2009)

The phase diagram and critical behavior of a simple toy model for DNA zipping/unzipping are examined in the framework of the Bethe approximation. The effects of solvent quality are included and found to lead to a variety of different thermodynamic behaviors.

DOI: [10.1103/PhysRevE.79.051108](https://doi.org/10.1103/PhysRevE.79.051108)

PACS number(s): 05.40.Fb, 05.20.-y, 05.50.+q, 36.20.-r

I. INTRODUCTION

Self-avoiding walk (SAW) models and their variants have been used for decades to gain insight into the physics of real polymer systems, with remarkable success [1]. Such models are now again of interest as toy models of biopolymers such as DNA and RNA molecules [2–6]. The effect of the base pairing is modeled by allowing the walk to visit the lattice bonds twice. In the context of the work presented here, the difference between DNA-like models and RNA-like models are the allowed walk configurations; DNA models allow at most four links of the walk to meet at a site (see Fig. 1), while RNA allow up to eight.

Recently several models of this type have been presented, either as models for biopolymers, or simply as interesting variants of restricted random walk models, in which different weightings are given to multiply visited sites or bonds [7–9]. What is missing from all these models are interactions representing the interactions with solvent molecules. In this paper we propose a step at filling this void.

In this paper, we introduce a variant of the lattice two-tolerant self-avoiding walk model [2,3,10] which mimics the zipping/unzipping of a DNA model. In the model introduced here we also include solvent effects by including an attractive interaction energy between nonconsecutively visited nearest-neighbor sites. In this paper we present an extended mean-field-type calculation of the phase diagram (the Bethe approximation). The phase diagram is found to be unexpectedly rich. In many DNA models there is a further restriction in that bases are only allowed to pair up if they are the same distance from one end along each of the two chains. This restriction is relaxed in the model presented here.

II. MODEL

The model studied in this paper consists of a noncrossing random walk on a square lattice limited to visit each bond and each site at most twice. The model is chosen to model the zipping/unzipping of a DNA molecule, thus the allowed configurations of the two-tolerant walk are further restricted such that a site may only be visited twice if one of the adjoining bonds is doubly visited. The allowed configurations are shown in Fig. 1.

Each segment of the walk represents a monomer (base) whereas doubly visited bonds represent a coarse-grained description of paired bases. The difference in affinity of the

DNA molecule with itself and with the solvent may be modeled by a solvent-mediated attractive interaction between nonconsecutive visited nearest-neighbor sites on the lattice. Solvent-mediated interactions carry an attractive energy $\epsilon_S < 0$ and doubly visited bonds yield an attractive energy $\epsilon < 0$.

The thermodynamic behavior may be investigated by introducing the grand-canonical partition function, \mathcal{Z} , from which many of the relevant thermodynamic quantities may be calculated. The grand-canonical partition function is given by

$$\mathcal{Z} = \sum_{\text{walks}} K^N \tau^{N_1} \gamma^{N_2}, \quad (1)$$

where N_1 is the number of solvent-mediated interactions, N_2 is the number of doubly visited bonds and $\tau = \exp(-\beta\epsilon_S)$, $\gamma = \exp(-\beta\epsilon)$, and $\beta = 1/kT$. The fugacity, which controls the average length of the walk, is denoted by K , and N is the total length of the walk. The two are related through

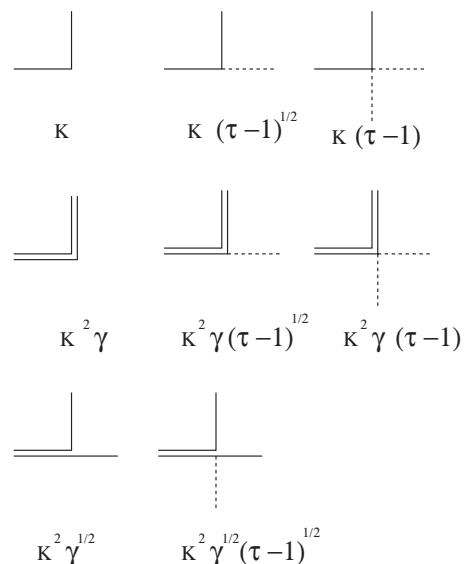


FIG. 1. Configurations at a lattice site on the square lattice. Polymer bonds are represented by solid lines while the solvent interactions are represented by dashed lines. The empty-site configuration is not shown here.

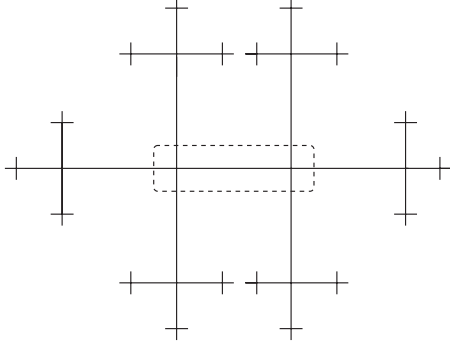


FIG. 2. The Bethe lattice representation of the two-dimensional lattice. The dotted box shows the central bond, exhibiting the desired square-lattice geometry.

$$\langle N \rangle = K \frac{\partial \ln \mathcal{Z}}{\partial K}. \quad (2)$$

The average length increases as K is increased.

III. BETHE APPROXIMATION

In this section we briefly describe the Bethe approximation. For a good discussion of the Bethe approximation see Ref. [11]. The model of interest is studied on the infinite Bethe lattice chosen to have the correct local geometry. The lattice chosen for the square lattice is shown in Fig. 2. The Bethe lattice is a hierarchical lattice built recursively from a central bond by adding k new bonds to each extremity. To each dangling bond we add k more bonds, and so on, such that no loops are formed. Due to the hierarchical nature of the lattice, it is possible to build up expressions for the partition function recursively. To see this, it is convenient to consider the lattice as being divided into two branches, left and right for the example shown in Fig. 2. We may introduce the partial partition functions W_σ^l and W_σ^r for the left- and right-hand branch, respectively. These partition functions are conditional upon the state σ of the central bond. In our model there are four possible states: (1) empty (state 0), (2) occupied with a link of the walk (state 1), (3) occupied with a solvent-mediated nearest-neighbor interaction (state S), and (4) occupied with a doubly visited bond (state 2).

By symmetry, the left and right branches will have the same partial partition functions, and so the l, r designation will be dropped. Each branch may be subdivided into k sub-branches, such that the W_σ may be expressed in terms of the partial partition functions of the sub-branches. This procedure may be continued until the boundary bonds are reached. In order to do this explicitly, it is convenient to introduce the notion of the “generation” of a link, n , which is simply the distance of the link from the boundary. As a concrete example, consider the calculation of $W_1^{(n)}$, the partial partition function conditional on the central bond being occupied by a link of the walk. We must consider all the configurations on the bonds of the generation $(n-1)$, of which there are three for the two-dimensional square lattice example shown in Fig. 2, which are compatible with the occupied central bond. Clearly there must be a bond leaving in one of the three

directions; the other two bonds may be empty or occupied by a solvent-mediated interaction. The weight $W_1^{(n)}$ is simply the sum of the Boltzmann weights corresponding to all these configurations, multiplied by the weight for adding the central link. To avoid the divergence of the partial partition functions it is convenient to introduce normalized partition functions $w_\sigma^{(n)} = W_\sigma^{(n)}/q_n$ [4] with q_n chosen such that

$$\sum_\sigma w_\sigma^{(n)} = 1. \quad (3)$$

This leads to recursion relations for the (normalized) partial partition functions:

$$w_\sigma^{(n)} = \frac{\lambda_\sigma}{q_n} \sum_{\{\sigma_i\}} C_{\sigma, \{\sigma_i\}} \prod_{i=1}^k w_{\sigma_i}^{(n-1)}, \quad (4)$$

where $\{\sigma_i\}$ is the set of states of the k links forming generation $n-1$, λ_σ is the Boltzmann weight of the bond added at generation n , and the factor $C_{\sigma, \{\sigma_i\}} = 1$ if the choice of the states $\{\sigma_i\}$ is compatible with the central state σ , and zero otherwise.

It is known that there is no phase transition on the infinite Bethe lattice, since the number of boundary sites grows too rapidly. However the recursion relations may be used in the center of the lattice as self-consistency equations for the two point mean-field theory for the corresponding square lattice. In this case, we assume we have translational invariance and drop the generational superscripts. The equilibrium states are then given by solutions of the following set of recursion relations:

$$w_0 = \frac{1}{q} \{w_0^3 + 3w_1^2(w_0 + w_S + w_2) + 3w_2^2(w_0 + w_S)\}, \quad (5)$$

$$w_1 = 3 \frac{K}{q} w_1(w_0 + w_S)(w_0 + w_S + 2w_2), \quad (6)$$

$$w_S = 3 \frac{(\tau-1)}{q} [w_1^2(w_0 + w_S + w_2) + w_2^2(w_0 + w_S)], \quad (7)$$

$$w_2 = 3 \frac{K^2 \gamma}{q} (w_0 + w_S) [w_1^2 + w_2(w_0 + w_S)], \quad (8)$$

$$q = w_0^3 + 3w_1^2(w_0 + w_S + w_2) + 3w_2^2(w_0 + w_S) + 3Kw_1(w_0 + w_S) \times (w_0 + w_S + 2w_2) + 3K^2\gamma(w_0 + w_S)[w_1^2 + w_2(w_0 + w_S)] + 3(\tau-1)[w_1^2(w_0 + w_S + w_2) + w_2^2(w_0 + w_S)]. \quad (9)$$

The partial partition functions give the contribution to one branch of the total partition function, the total (normalized) partition function conditional upon the state of the central bond is then given by the product of the weights for the left and right branches. Each of the partial partition functions includes the Boltzmann weight corresponding to the state of the central bond, which is thus counted twice in the full partition function. This double counting is corrected by dividing each term by the relevant Boltzmann weight. Summing over all the possible states for the central bond gives the total (normalized) partition function, z :

$$z = \sum_{\sigma} \frac{w_{\sigma}^2}{\lambda_{\sigma}}. \quad (10)$$

In the usual way, the probability of finding a given bond in state σ is given by the partition function conditioned upon this state divided by the total partition function, i.e.,

$$p_{\sigma} = \frac{w_{\sigma}^2}{z\lambda_{\sigma}}. \quad (11)$$

It should be noted that the density ρ of the walk on the lattice is simply

$$\rho = p_1 + 2p_2. \quad (12)$$

Another quantity of interest is the fraction of paired segments given by

$$\Phi = \frac{2p_2}{\rho}. \quad (13)$$

The grand potential per site may be related to z and q through the relation

$$\beta f = \frac{(k-1)\ln z - 2\ln q}{2} = \ln z - \ln q, \quad (14)$$

for the square lattice ($k=3$). For a full derivation of this expression see [4]. When multiple solutions to the recurrence relations exist, the solution with the lowest value of the grand potential is the stable equilibrium solution.

IV. RESULTS

It is convenient to recast recursion relations (5)–(8) by setting

$$x_1 = \frac{w_1}{w_0},$$

$$x_S = \frac{w_S}{w_0},$$

$$x_2 = \frac{w_2}{w_0}.$$

The different phases and transition lines correspond to different solutions of the recursion equations. These equations have a trivial solution $x_1=x_S=x_2=0$ corresponding to the finite polymer phase. The density is trivially zero, since the lattice is infinite. In what follows we refer to this phase as the O phase. Nontrivial solutions may be found by setting different values of the parameters K , τ , and x_S and then solving numerically for γ , x_1 , and x_2 .

The trivial zero-density phase is separated from a region where the walk fills the lattice with a finite density by a surface $K=K_{\infty}(\tau, \gamma)$ where the average length of the walk first diverges. In the finite-density region we find two dense phases: a phase (I) in which a finite fraction of the segments of the walk are paired, i.e., it is characterized by $0 < \rho < 1$ and $0 < \Phi < 1$ and a fully paired dense phase (II) in which

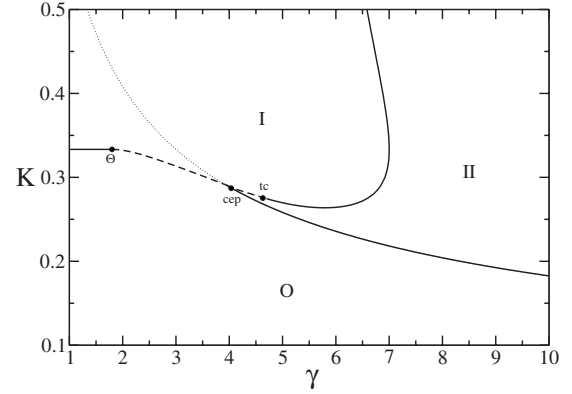


FIG. 3. The phase diagram in the K - γ plane, with $\tau=1$. O denotes the zero density phase, I the ordinary dense phase, and II the fully paired dense phase. Dashed and solid lines denote first- and second-order transitions, respectively. The tricritical Θ point on the O-I transition line is given by $K=1/3$ and $\gamma_{\Theta}=9/5$. The second-order transition between phases O and II is given by the line $3K^2\gamma=1$ and finishes in a critical end point. The dotted line is the extension of this line and is included to guide the eye.

every segment is paired, i.e., it is characterized by $0 < \rho < 1$ and $\Phi=1$.

When $\gamma=0$ (giving $x_2 \equiv 0$) the pairing of segments is forbidden and the model corresponds to the standard interacting self-avoiding walk model. For this model, also called the Θ -point model [1], phases O and I are separated by a critical transition line for small enough τ and a first-order line for large enough τ . These two behaviors are separated by the tricritical Θ point. This tricritical point extends to a line of tricritical points as γ is allowed to increase, separating a critical region in the SAW universality class from the first-order region, where the walk fills the lattice with a finite density. It is possible to determine analytically the region occupied by the SAW phase, and so the equation of the tricritical line, as follows: from phase I ($x_1 \neq 0$) we take the limit $x_S \rightarrow 0$ and $x_2 \rightarrow 0$ in order to approach the boundary with phase O. Since the transition is continuous, this limit leads to $x_1 \rightarrow 0$ and gives $K_{\infty}=1/(2d-1)=1/3$. The tricritical point is obtained by allowing the parameters x_1 , x_S , and x_2 to be vanishingly small but nonzero along the transition line $K=1/3$. We obtain the tricritical condition:

$$\gamma_{\Theta}(\tau) = \frac{9(3-2\tau)}{4(\tau-1) + 5(3-2\tau)}. \quad (15)$$

It is important to note that there is no guarantee that the extended line is in the same universality class as the Θ point, but we will nevertheless use the Θ subscript to differentiate it from other special points in the phase diagram. We shall return to this point later. If the pairing interaction is taken to be attractive, then $\gamma \geq 1$, this leads to the result that the transition line O-I becomes fully first order for $\tau > 4/3$.

The phase diagram for $\tau=1$ is shown in Fig. 3, clearly showing the location of the tricritical Θ -like point, found to be $K=1/3$ and $\gamma_{\Theta}=9/5$. The phase boundary of phase I is smooth, with the boundary between phases I and II being partly first order and partly second order. This gives rise to

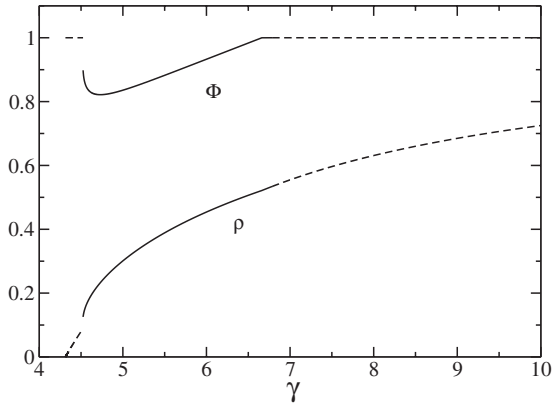


FIG. 4. The two order parameters, the density ρ and the fraction of paired segments Φ as a function of γ for $\tau=1$ and $K=0.278$. Solid and dashed lines denote their values in phases I and II.

another tricritical point at $\gamma_{tc} \approx 4.7$. The transition line between phases O and II is found to be continuous, and the equation for this line may be determined exactly: from phase II ($x_1=0$ and $x_2 \neq 0$) we take the limit $x_S \rightarrow 0$ to approach the boundary with phase O. Using the fact that the transition is continuous, we determine that $x_2 \rightarrow 0$ leads to the condition $3K^2\gamma=1$. This line terminates in a critical end point at $\gamma_{cep} \approx 3.94$.

For $\tau=1$, i.e., in absence of solvent effects, the phase diagram is similar to the phase diagram obtained by Pretti [4] for an RNA-like model with nonzero stacking energy, except that in Pretti's case the equivalent to the O-II line was found to be first order. The stacking energy had the effect of favoring the absence of multiple double bonds meeting at a site, and so made the model more like the model presented here.

Before examining other values of τ , it is of interest to discuss the different phases shown in Fig. 3 and the relevant order parameters. Two parameters are of interest for differentiating the different phases: the density of occupied lattice bonds ρ and the fraction of paired segments Φ . These two quantities are plotted in Fig. 4 for $K=0.278$, so as to pass through each phase. In phase O $\rho=0$ trivially since the walk is finite on an infinite lattice. In the framework of the Bethe approximation it is not possible to calculate Φ in this phase ($\rho=0, p_2=0$). The value of K is chosen to cross the transition from phase O to phase II under the critical end point. We clearly see that ρ increases smoothly from 0, indicative of a critical transition, and that $\Phi=1$ indicating that phase II corresponds to a saturated doubly occupied phase. Again we remind the reader that the apparent jump in Φ is simply related to the fact that it is not possible to calculate Φ in phase O (where $p_1=p_2=0$). The value of K is chosen to cut back into phase I between the critical end point and the second tricritical point. We can clearly see in Fig. 4 that Φ drops discontinuously from 1 and there is a small jump in ρ . The transition from phase I to phase II is again continuous and the interest of Φ as order parameter is clearly seen; there is no evidence of a phase transition in ρ , which is not a good order parameter for phase II.

As τ is increased, the phase diagram changes. The first change is that the Θ -like tricritical transition is pushed out of the domain of interest ($\epsilon > 0$ or $\gamma > 1$). This occurs, as pre-

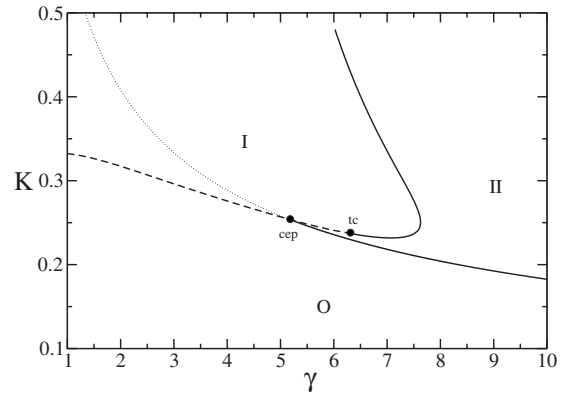


FIG. 5. The phase diagram in the $K-\gamma$ plane, with $4/3 < \tau = 1.4 < 3/2$. Dashed and solid lines denote first- and second-order transitions, respectively. The point marked cep is a critical end point, while tc indicates a tricritical point. The dotted line corresponds to the line $3K^2\gamma=1$.

viously stated, for $\tau=4/3$. This case is shown in Fig. 5. The calculation for the transition line between phases O and II remains valid as τ is increased, but for large enough τ the transition is found to be first order, see the phase diagrams in Figs. 6 and 7. We may determine the condition for this change in behavior by allowing the parameters x_S and x_2 to be vanishingly small but nonzero along the transition line. We obtain the condition $\tau=3/2$. Thus the O-II line is second order and given by the equation $3K^2\gamma=1$ for $\tau \leq 3/2$ and it is first order for $\tau > 3/2$, and falls below the line $3K^2\gamma=1$. This behavior may be understood in another way; as $\gamma \rightarrow \infty$, the walk becomes a doubled up self-avoiding walk, with an effective fugacity per bond $\tilde{K}=K^2\gamma$. We may then apply the known results to find that for $\tilde{K} < 1/3$ the transition from phase O to the dense phase is continuous, and for $\tilde{K} > 1/3$ the transition is first order. There is a Θ transition when $\tau=3/2$ and $3\tilde{K}=3K^2\gamma=1$. Since phase II is a saturated phase, the walk remains a doubled up self-avoiding walk for finite γ , hence the observed result, and the prediction that for τ

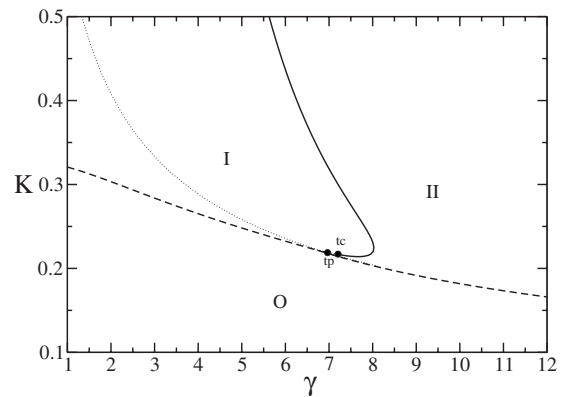


FIG. 6. The phase diagram in the $K-\gamma$ plane, with $3/2 < \tau = 1.6 < 1.75$. Dashed and solid lines denote first- and second-order transitions, respectively. The point marked tp is a triple point, while tc appears to be a tricritical point. The dotted line corresponds to the line $3K^2\gamma=1$.

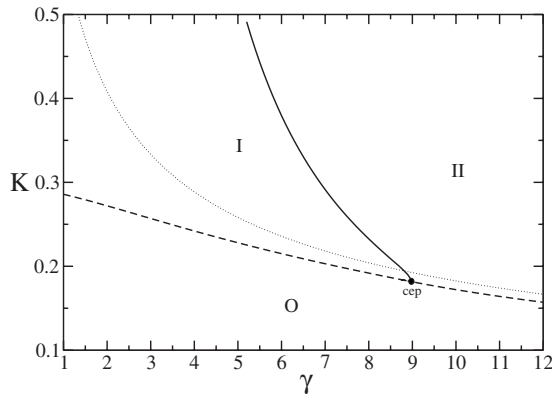


FIG. 7. The phase diagram in the $K-\gamma$ plane, with $\tau=2 > 1.75$. Dashed and solid lines denote first- and second-order transitions, respectively. The dotted line corresponds to the line $3K^2\gamma=1$.

$=3/2$ the transition line between phase O and phase II is of the same type as the Θ transition. It would be interesting to ascertain whether this prediction is maintained if a full calculation for the model is performed. For $\tau < 3/2$ (see Fig. 5) the first-order line between phases O and I and the first-order line between phases I and II are tangential to each other at the critical end point which defines the end of the critical line between phases O and II. However, for $3/2 < \tau < 1.75$ (see Fig. 6) the three first-order lines meet at a triple point (however the first-order lines between phases O and I and between I and II are still tangential). The transition line between phases I and II changes from first order to second order at a tricritical point.

Another interesting change arises when $\tau \approx 1.75$ (see Fig. 7). The phase transition between phases I and II becomes totally second order, and the first-order boundary to phase O becomes smooth, and the point where the three phases meet is again a critical end point.

Of particular interest is the phase diagram in the plane $K_\infty(\gamma, \tau)$, where the average length of the walk first diverges. This plane is particularly important, since it corresponds to the phase diagram which is seen in a canonical simulation of single finite-sized walk in the limit of infinite walk length. Indeed it is sometimes referred to as the “thermodynamic

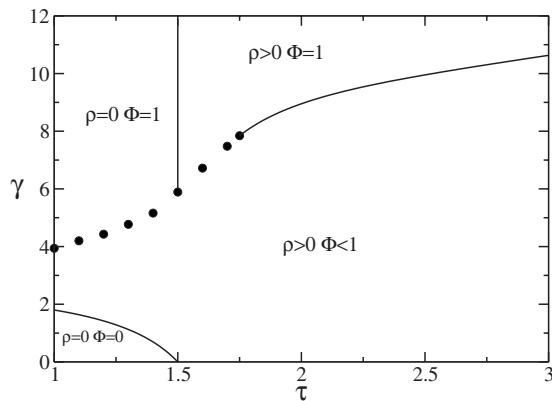


FIG. 8. Phase diagram in the $K_\infty(\gamma, \tau)$ plane. Solid lines represent continuous phase transitions, while the points indicate a first-order transition line.

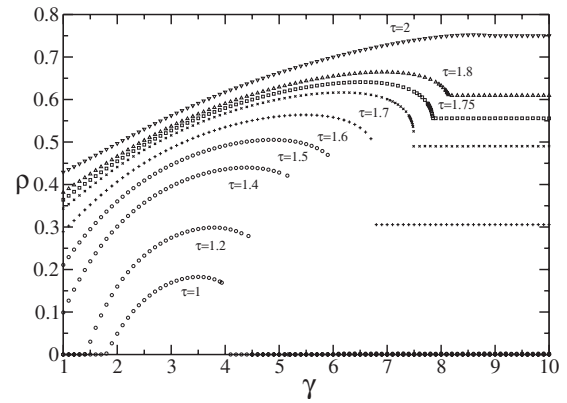


FIG. 9. Density ρ plotted in the $K_\infty(\gamma, \tau)$ plane as a function of γ for different values of τ .

limit” for the canonical dilute polymer problem. The phase diagram in the K_∞ plane is shown in Fig. 8. There are four different phases differentiated by the values of ρ and Φ . There are two critical $\rho=0$ phases, both of which should be in the self-avoiding walk universality class. One has $\Phi=0$ (the standard SAW phase) while the other has $\Phi=1$ corresponding to a double SAW phase. There are two noncritical phases, one with $\Phi=1$, corresponding to a standard type of collapsed phase as seen in the Θ -point model, except that the walk is doubled. The other phase has a value of Φ different from both 0 and 1.

It is interesting to note that $\gamma_\Theta(\tau)$ decreases as τ increases, making it easier for the walk to pass from the $\rho=0, \Phi=0$ phase to the neighboring dense phase as τ is increased. At first sight this seems odd, however, the formation of some double bonds enables the formation of a branched polymer type conformation, which is more naturally space filling. This would lead one to expect that perhaps the phase boundary between the $\rho=0, \Phi=0$ phase and the $\rho>0, 0 < \Phi < 1$ phase may not be in the Θ -point universality class. The mean-field nature of the Bethe approximation does not enable one to investigate this point.

From previous discussion, the line separating the two $\Phi=1$ phases, found at $\tau=3/2$, may be expected to be a line of tricritical points in the Θ -point universality class. This line appears to end at a tricritical end point on a first-order line

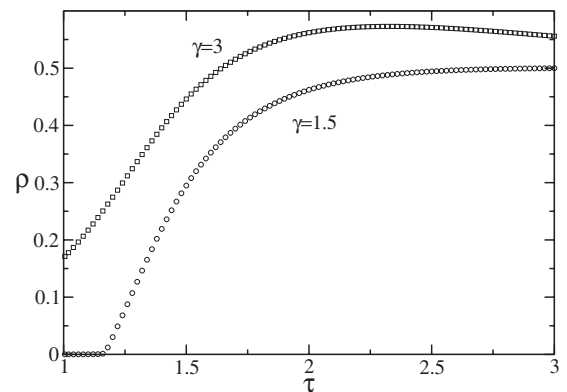


FIG. 10. Density ρ plotted in the $K_\infty(\gamma, \tau)$ plane as a function of τ for different values of γ .

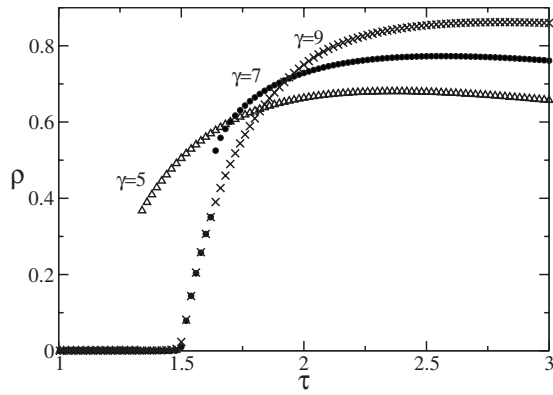


FIG. 11. Density ρ plotted in the $K_\infty(\gamma, \tau)$ plane as a function of τ for different values of γ .

($\tau=3/2, \gamma \approx 5.89$), which in its turn ends at another tricritical point at $\tau \approx 1.75, \gamma \approx 7.845$. While much could change once fluctuations are included, one could expect much of the behavior seen here to remain in the real system. Figures 9–12 show different plots of ρ and Φ showing their behavior at the different phase boundaries, confirming the orders of the transitions shown. In Fig. 9 the density is plotted as a function of γ for different values of τ . For small values of τ two transitions may be clearly seen: The first is a second-order transition where the density becomes nonzero continuously and the second, at higher values of γ , is first order, where the density becomes discontinuously zero again. At higher values of τ , the first transition no longer exists for attractive pairing energies ($\gamma > 1$). For $\tau > 3/2$ the density at the second transition jumps to a nonzero value, discontinuously at first, but continuously for higher γ (γ bigger than about 1.75). Figures 10 and 11 show plots of ρ as a function of τ for different γ and Fig. 12 shows Φ , the proportion of paired

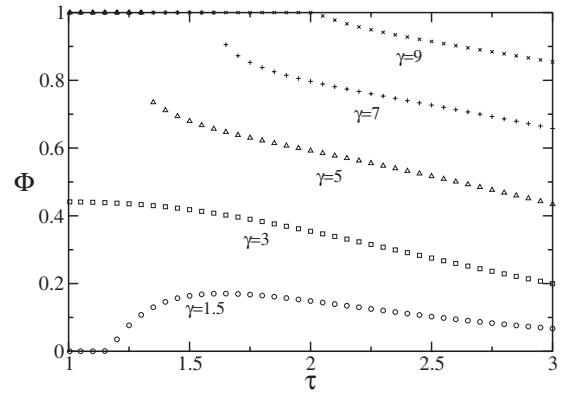


FIG. 12. Fraction of paired bonds Φ plotted in the $K_\infty(\gamma, \tau)$ plane as a function of τ for different values of γ .

bonds, for the same values of γ . As may be seen by comparing these figures, ρ and Φ are good order parameters for different phase transitions, notably, when $\gamma=9$ the density plot does not show the presence of the phase transition at $\tau \approx 2$, clearly shown in the plot of Φ .

V. DISCUSSION

In this paper we have investigated the phase diagram of a noncrossing random walk model on the square lattice where the walk is allowed to visit lattice bonds twice, and the site configurations are chosen to represent the zipping/unzipping of a DNA molecule. Solvent quality is included through attractive nearest-neighbor interactions. This model is examined in the framework of the Bethe approximation and found to have a rich phase diagram. While there are limitations to the method, it usually captures the essential features of the phase diagram [12,13]. It would be of interest to look at this model using some other method to confirm and examine further the phase diagram presented.

-
- [1] C. Vanderzande, *Lattice Models of Polymers* (Cambridge University Press, Cambridge, England, 1998).
- [2] M. Baiesi, E. Orlandini, and A. L. Stella, *Phys. Rev. Lett.* **91**, 198102 (2003).
- [3] P. Leoni and C. Vanderzande, *Phys. Rev. E* **68**, 051904 (2003).
- [4] M. Pretti, *Phys. Rev. E* **74**, 051803 (2006).
- [5] R. A. Zara and M. Pretti, *Physica A* **371**, 88 (2006).
- [6] R. A. Zara and M. Pretti, *J. Chem. Phys.* **127**, 184902 (2007).
- [7] J. Krawczyk, T. Prellberg, A. L. Owczarek, and A. Rechnitzer, *Phys. Rev. Lett.* **96**, 240603 (2006).
- [8] P. Serra and J. F. Stilck, *Phys. Rev. E* **75**, 011130 (2007).
- [9] T. J. Oliveira, J. F. Stilck, and P. Serra, *Phys. Rev. E* **77**, 041103 (2008).
- [10] E. Orlandini, F. Seno, A. L. Stella, and M. C. Tesi, *Phys. Rev. Lett.* **68**, 488 (1992).
- [11] R. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982).
- [12] S. Lise, A. Maritan, and A. Pelizzola, *Phys. Rev. E* **58**, R5241 (1998).
- [13] D. P. Foster and M. Aniambossou, *Phys. Rev. E* **77**, 061121 (2008).