Theory of DNA melting based on the Peyrard-Bishop model

Yong-li Zhang,[1] Wei-Mou Zheng,[1] Ji-Xing Liu,[1] and Y. Z. Chen[1,2]
[1]*Institute of Theoretical Physics, Academia Sinica, P.O. Box 2735, Beijing 100080, China*
[2]*Department of Physics, Purdue University, West Lafayette, Indiana 47907-1396*
(Received 13 March 1997)

The DNA melting based on the Peyrard-Bishop (PB) model is systematically investigated. Our study on the eigenvalues and eigenvectors of the transfer integral equation for the original PB model points out that the eigenvectors are composed of two kinds: discrete bound states that constitute the internal states of a DNA molecule and the continuous unbound states that represent its dissociated states. Another process controlling the melting of DNA—the dissociation equilibrium between duplex and single-stranded DNAs—is introduced, which leads to an extended model applicable for a realistic DNA chain with a finite number of base pairs. Based on the expansion of kernels, the calculations of the thermodynamic quantities of the system are reduced to multiplication of matrix series. Calculations on model block DNAs show the method is much more efficient than molecular dynamic simulation and has enough high precision to handle the melting of a natural DNA with arbitrary sequence. The discreteness effect and nonlinear effect of the model are discussed based on the Gaussian model. Rigorous melting curves for periodic DNA with two and three base pairs in a unit cell and boundary effects are presented by the transfer integral approach. [S1063-651X(97)01012-X]

PACS number(s): 87.10.+e, 63.70.+h, 64.70.−p

I. INTRODUCTION

The structure and dynamics of DNA are the key to understanding its biological effects and have long been subjected to extensive theoretical studies. Different phenomena and properties of DNA have been studied theoretically including the melting (or denaturation) of DNAs [1], the large amplitude fluctuations or nonlinear excitations [2], the structural transitions such as *B-A* or *B-Z* transition [3], the stabilities of DNA complexes such as drug DNAs and protein DNAs [4], the interactions with molecules in surroundings such as water and ions [5], the structures of supercoiled DNAs [6], etc. Among these, the study on the melting of DNA plays an extremely important role in understanding the structure and dynamics of DNA.

However, so far, to our knowledge, there still has been no theory satisfactory enough to describe multiple interesting features of DNA melting. The reasons lie in several aspects. As a system with many degrees of freedom, it has rather different characteristics from those often encountered in condensed matter physics. Firstly, the structure and dynamics of DNA are very complicated [7]. Experiments and theoretical analyses indicate that many different excitations could possibly coexist in DNA. The motions of the sugar-phosphate backbones can be described by phonons [8], while the stretch of hydrogen bonds and the rotation of base pairs may pertain to solitons with completely different forms such as kinks, breathers, and probably other kinds of solitons [2]. Furthermore, these excitations are strongly coupled to each other and to the surrounding media [9]. It was pointed out that a good DNA model must be nonlinear in nature [10]. Secondly, the interactions that govern the structures and dynamics of DNA are not very clear. Because of the complexity of biomolecules, we often have to understand their properties at the level of atomic groups. The interactions between them are usually many body in nature and cannot be simplified as the central two-body potentials. Finally, biomolecules exist in and interact with the environment, and thus often cannot be studied individually because of their strong couplings. A biomolecule, which can be regarded as a thermodynamic system [11], has a large surface to volume ratio compared with the usual thermodynamic systems. This is similar to surface physics, where the materials absorbed on a surface can cause great structural changes, such as reconstruction. Water molecules and ions also have a significant effect on the structure of biomolecules. The dynamic interaction between biomolecules and water, the hydrophobic interaction, plays an important role in the structure and dynamics of biomolecules. These interactions, however, remain to be understood.

A variety of means for simplification have been adopted in the theories on the dynamics of DNA. They often fall into one of two categories. One is to simplify its structure or states of motion, the other is to simplify the interaction potentials. At present, there are two principal theories on the melting of DNA, which are both good examples of the above categories. One is the helix-coil transition theory based on the Ising model introduced in late 1950s [12]. In this model, the motion of each base pair is assumed to be in only two states: hydrogen bonded (intact) or non-hydrogen bonded(open). The other is the lattice dynamic theory based on the modified self-consistent phonon approach (MSPA) [13]. This theory introduced in 1984 by Prohofsky and co-workers has been widely applied to investigate the melting of DNA and the interactions between DNA and drugs. In their model they employed the real configurations of DNAs given by experiment, but assumed harmonic potentials for all bonds except hydrogen bonds, which were represented by Morse potentials with appropriate parameters. Even with such a simplification of the interactions, it still seems beyond our capacity to find rigorous treatment. In their calculation, the Morse potentials were actually replaced by harmonic potentials with force constants self-consistently determined according to the principle of minimization of free energy. Thus

it is a linear lattice dynamic theory in nature.

To some extent, the model adopted in this paper is a combination of the above two DNA melting theories. The model was presented by Peyrard and Bishop in 1989 [10]. In the PB model, each backbone of DNA is simplified as a one dimensional chain with nearest neighboring potentials representing stacking energy. The potential of the hydrogen bond in base pair is assumed to be Morse potential between the units of two chains. The main advantages of this model lie in two aspects. On one hand, for the homogeneous chain, the model can be solved rigorously using the technique of transfer integral (TI). By comparing the rigorous results with those obtained from perturbative methods such as self-consistent phonon approach (SCPA), one can investigate the nonlinear nature of the DNA melting [14]. On the other hand, it is a unified model treating the DNA melting and the large amplitude fluctuation of base pairs in DNA. Dauxois and Peyraud have widely investigated the nonlinear excitations of this model system [15].

But as a model for DNA melting, it has only been successfully applied to a special case where the base-pair composition is homogeneous and the chain is infinite in length. How to extend the model to a more realistic one, i.e., having arbitrary sequences and open boundaries, is still an open question. There are two main difficulties in making this extension. One is related to the divergence of the thermodynamic quantities of this model system [13,16]. It has been shown that the partition function of the Peyraud-Bishop (PB) model is convergent only in the limit of infinite number of base pairs N. The other is the inhomogeneity of the base sequence where the transfer integral technique is no longer valid in this case. In this paper we discuss the methods for solving these difficulties.

We will show that the first difficulty can be solved by regarding the melting of DNA as being governed by two processes: the internal unwinding motion within a single duplex DNA, which can be described by the PB model, and the dissociation equilibrium between double-stranded DNA (C_2) and single-stranded DNA (C_1), i.e.,

$$C_2 \rightleftharpoons 2C_1,$$

which can be well described by the law of mass action and has been included in the Ising model of DNA melting [1]. Here the scope of the ''internal'' and the ''external'' motion is approximately separated by the *dynamic diameter d* in phase space. With the separation of these two different motions, the divergence of the thermodynamic quantities of the PB model is avoided.

The second problem can be solved by the expansions of kernels with an appropriate set of orthonormal bases. The partition function can then be expressed as a series of products of matrices, which are related to the transfer matrices in the Ising model. So the continuum model can be well approximated by an extended Ising model with M-component Ising spins, where M is the number of base functions used to expand the kernels. Thus we extend the two-state helix-coil transition theory that relies on macroscopic parameters to M-state theory, which in principle has only microscopic parameters. Compared with Prohofsky's lattice dynamic theory, this model can be viewed as a simplified lattice model that only retains the basic lattice structure and core degrees of freedom in a DNA chain, but emphasizes the nonlinear effect hidden in it.

This paper is organized as follows: In Sec. II, after introducing the PB model, we investigate the technique of transfer integral with a highly singular homogeneous integral equation, which has not been encountered in the standard TI technique and well studied in mathematics as we know. We exhibits the characteristics of the eigenvalues and eigenvectors of the TI equation resulted from this feature. Then we present the dissociation process of duplex DNAs to single-stranded DNAs so as to construct a complete theory of DNA melting according to the partition of the states of DNA molecule. In Sec. III, we investigate the discreteness effect and nonlinear effect of the model, which are two main features of the system, based on the rigorous results of the Gaussian model. Sections IV and V can be considered as the further applications of the TI technique. In Sec. IV, we study the boundary effect of a homogeneous DNA with finite length while in Sec. V we calculated the melting curves of periodic DNAs with two or three base pairs in a unit cell. We find there are no fine structures in these cases. In Sec. VI, we give two versions of our *extended transfer* matrix approaches and compute the melting profiles of two block DNAs with different block sizes. The calculations show our approach is practical and has the precision high enough to compute the melting profile of a natural DNA with arbitrary sequence. Section VII is a discussion.

II. MODEL AND ITS STATISTICAL MECHANICAL BASIS

A. Peyrard-Bishop model

The detailed description of the PB model can be referred to Refs. [14,17,19]. Here we only give a brief introduction. The motion of the model system can be described by two kinds of variables: the displacement of the center of mass of each base pair and the separation of the bases in the same base pair, denoted by y_n. It is the latter that determines the melting of DNA. After the decoupling of the two kinds of variables, the concerned Hamiltonian of the system is

$$H_y = \sum_n \left[\frac{1}{2} m \dot{y}_n^2 + w(y_n, y_{n-1}) + V(y_n) \right]. \tag{1}$$

Here $V(y)$ is Morse potential

$$V(y) = D(e^{-ay} - 1)^2 \tag{2}$$

and $w(y_n, y_{n-1})$ the potential between the nearest neighboring base pairs, representing stacking energy. This Hamiltonian is similar to the Hamiltonians used in the models of structural phase transitions except that the on-site potentials there have two or three minima [20,21]. Therefore results from the theories on structural phase transitions can be used to help the understanding of DNA melting as shown later.

The form of $w(y_n, y_{n-1})$ is crucial to the model. At present its direct experimental determination or theoretical calculation is still unavailable. As a result empirical forms have to be developed based on some general principles. In the first version of the PB model, they adopted a harmonic potential. It has been shown that the simple harmonic form

cannot give a good quantitative description of DNA melting because the melting scopes given are too wide to determine a melting temperature accurately. Dauxois and Peyraud presented an anharmonic potential to describe the stacking energy later [17], that is,

$$w(y_n,y_{n-1})=\frac{k}{2}(1+\rho e^{-\alpha(y_n+y_{n-1})})(y_n-y_{n-1})^2. \quad (3)$$

This modified version of the PB model was proven to be successful because it may give not only a qualitative but also a quantitative description of DNA melting. It is to be our focus in this paper.

B. Technique of transfer integral

Let

$$K(x,y)=\exp(-\beta\{w(x,y)+\tfrac{1}{2}[V(x)+V(y)]\}), \quad (4)$$

where the Boltzmann factor is $\beta=1/kT$. It is evident that

$$K(x,y)=K(y,x). \quad (5)$$

The partition function of the system described by H_y can then be written as

$$Z=\int dy_1 dy_2\cdots dy_N K(y_1,y_2)K(y_2,y_3)\cdots K(y_N,y_1) \quad (6)$$

in which a periodic condition is adopted and a trivial momenta factor is omitted.

In order to solve the partition function, one introduces the integral equation

$$\int K(x,y)\varphi(y)dy=\lambda\varphi(x). \quad (7)$$

Because of the symmetry of the kernel and

$$K(x,y)>0,$$

if we further assume

$$||K(x,y)||\equiv\left(\int\int [K(x,y)]^2 dxdy\right)^{1/2}<\infty, \quad (8)$$

the integral equation then contains a positive Hilbert-Schmidt type kernel [18,22]. So it has a set of positive eigenvalues and orthonormal eigenvectors. If we denote the eigenvalues as λ_1, λ_2, etc. in descending order, and $\varphi_1(x),\varphi_2(x)$, etc. as the corresponding eigenvectors, then

$$\int dx\varphi_n(x)\varphi_m(x)=\delta_{nm}, \quad (9)$$

and

$$\sum_{n=1}^{+\infty}\varphi_n(x)\varphi_n(y)=\delta(x-y). \quad (10)$$

$K(x,y)$ can be expanded as [18]

$$K(x,y)=\sum_n \lambda_n\varphi_n(x)\varphi_n(y). \quad (11)$$

By substitution of the expansion expression of $K(x,y)$ into Eq. (6) and applying the orthonormal and complete relations (9) and (10), we have

$$Z=\sum_{n=1}^{+\infty}\lambda_n^N. \quad (12)$$

Similarly, from transform invariance of the system, we obtain an average thermal expansion of each base pair,

$$\langle y\rangle=\frac{1}{Z}\sum_n \langle n|y|n\rangle\lambda_n^N, \quad (13)$$

where

$$\langle n|y|m\rangle\equiv\int \varphi_n(y)y\varphi_m(y).$$

At the thermodynamic limit, Eqs. (12) and (13) can be simplified as

$$Z=\lambda_1^N \quad (14)$$

and

$$\langle y\rangle=\langle 1|y|1\rangle, \quad (15)$$

respectively.

In the above we have given the standard TI technique widely used in one- or quasi-one-dimensional systems [20,21,23]. However, significant differences exist between the PB model and those considered by the standard TI technique. In the models using the standard TI technique, all on-site potentials are unbound, thus the condition (8) exists. However, the Morse potential in the PB model is bound [13], which implies formula (8) does not hold. So the kernel (4) is not a Hilbert-Schmidt type one but a singular kernel defined on the space $[-\infty,+\infty;-\infty,+\infty]$. It is this characteristics of the kernel that leads to the divergence of the partition function of the PB model. Because of the nonexistence of the prerequisite (8), we can easily see that the expansion of the kernel, which is a crucial step for the TI technique, does not hold any more. This indicates that at least in order to to carry out the TI technique, an upper bound for y_n needs to be set up. But this is not equivalent to saying that the integral equation (7) does not constitute an eigenvalue problem. In the following we first study the general features of the eigenvalues and eigenvectors of this TI equation based on numerical computation. Our strategy is to limit the kernel on a finite space $[a, b; a, b]$, therefore its norm exists, and then investigate the case where $a\to-\infty$ and $b\to+\infty$.

C. Eigenvalues and eigenvectors of the TI equation

A numerical solution of the integral equation is necessary for this statistical mechanical model. Two kinds of methods have been used by Dauxois and Peyrard to solve this TI equation [16]. One is to discretize the integral by means of summation formulas, then the problem is equivalent to find-

ing the eigenvalues and eigenvectors of a symmetric matrix. Here the choice of the quadrature rule is very important. They tried three different formulas: the trapezoidal rules, the Simpson rule, and the Bode's integration rules of order 6 and 10. The last one renders the best accuracy. But even for this one the needed dimension of the matrix should be as large as 1441×1441. The other approach is Kellog's method, which is convenient to compute the first several eigenvalues and eigenvectors. In this paper, we adopt the first kind of method but use Gauss-Legendre quadrature formula. We choose the order of the Gauss-Legendre node to be equal to the number of the points used to discretize the integration, therefore also equal to the dimension of the matrix [24]. With this modification, we find that a matrix with dimension as small as 70×70 is enough to give rather accurate results [25]. This faster algorithm is the guarantee of our work in the next sections.

At first, we find the relationship between eigenvalues and eigenvectors of the TI equation and the lower limit a is rather trivial. When $a<-0.3$ Å, they no longer change. So in the following we only concentrate on their dependences on the upper limit b.

Figure 1 shows the changes of the first several eigenvectors and eigenvalues with b and temperature. We can see that there are two different kinds of eigenvectors. One does not change with the value of b, indicating that it corresponds to a bound state; the other changes drastically, indicating that it pertains to an unbound state. When the temperature rises, the bound states gradually become unbound as can be seen from the comparison of Figs. 1(a) and 1(b). From Fig. 1(b) it can also be found in the interested temperature region from 80 K to critical point 350 K, there is only one bound state. As the temperature is so high that the last bound state becomes unbound, $\langle y\rangle$ will increase sharply, indicating the melting of the system. From Fig. 1(c), it can be seen that above the melting temperature, no bound state exists. Figure 1(d) shows the changes of the four biggest eigenvalues versus b. We find that for a bound state, its eigenvalue does not change with b, which is in agreement with the relationship of its eigenvector and b. While for unbound states, their eigenvalues become degenerate with the rise of b. This feature as well as the change of the unbound state with b indicates that all unbound states are degenerated and constitute continuous states. The discrete appearance of their eigenvalues is caused only by the finite value chosen for b. Here we can find that the spectrum of this eigensystem is similar to that derived from the Schrödinger equation describing a particle moving in a Morse potential. Because of this similarity in the mathematical structure of the two problems, we can see from Eqs. (12) and Eq. (13) that the melting of DNA can be compared to the dissociation of a two-atom molecule or the ionization of an atom under a substitution

$$\lambda_n = e^{-\beta\epsilon_n/N}, \tag{16}$$

where ϵ_n corresponds the energy of a molecule or an atom [26,27].

Figure 2 shows that the first bound state becomes wide when raising the temperature, but at their peaks they stay near zero. From Eq. (15), we find that the widths of these states actually represent the scopes of fluctuations of the base pairs at the corresponding temperatures. Noting their sharp peaks, we can easily see why a high order of the Gauss-Legendre node is needed.

Here we understand that, although the kernel is singular when $b\to+\infty$, due to the existence of the bound states, the TI technique still holds for the limit $N\to\infty$ if we get the thermodynamic limit by taking $N\to\infty$ before $b\to\infty$. From Eq. (13), we can see

$$\lim_{b\to\infty}\lim_{N\to\infty}\langle y\rangle = \langle 1|y|1\rangle. \tag{17}$$

This suggests that the extreme situation does exist. However, when N is finite, because of the existence of the unbound states in Eq. (13), $\langle y\rangle$ unavoidably diverges when $b\to\infty$, as can be seen from Fig. 3. Similar results were also obtained by Prohofsky and co-workers from molecular dynamics simulation [13]. This exhibits the inability of the PB model to describe the whole process of the DNA melting.

D. Chain dissociation and $\theta_{\rm ext}$

We can see that $b\to+\infty$ implies a single double-stranded DNA molecule occupying infinite space. Obviously it is not a real situation. In the solution of DNA molecules, a great deal of them are limited in a finite space. Every DNA molecule is actually in a finite effective space, thus setting an upper bound for b.

It also can be shown that when N is finite, the PB model cannot describe the melting completely. Let consider an extreme case where a DNA molecule has only one or several base-pairs. We find the melting of the duplex is reduced to a chemical reaction represented by formula $C_2\rightleftharpoons 2C_1$. This suggests that we can also regard the melting of DNA with bigger finite base pairs as a chemical reaction except this reaction takes place at a definite small temperature region only. Because the total numbers of particles are not conserved, the system should be described by a grand canonical ensemble, which usually leads to a complex theory. However, when the concentration is low, which is the very case of DNA solution in the melting experiments, the interactions between the solute molecules, except for the collisions, are negligible. Thus the system comprises an ideal solution. Suppose the solution constitutes of N_1 single-stranded DNA molecules and N_2 double-stranded DNA (duplex DNA) molecules. The grand partition function for the system is

$$\Xi(N_T,V,T) = \sum_{N_1=0}^{N_T} \frac{Z_1^{N_1}}{N_1!}\frac{Z_2^{N_2}}{N_2!}, \tag{18}$$

where Z_1 and Z_2 are the single-particle partition functions of single-stranded DNA and double-stranded DNA, respectively, and $N_T=N_1+2N_2$ is a constant, V the volume of the solution. Taking the maximum term in the summation for approximation, we obtain the law of mass action for the equilibrium $C_2\rightleftharpoons 2C_1$, i.e. [12],

$$\frac{Z_2}{Z_1^2} = \frac{N_2}{N_1^2} = K_{\rm eq}. \tag{19}$$

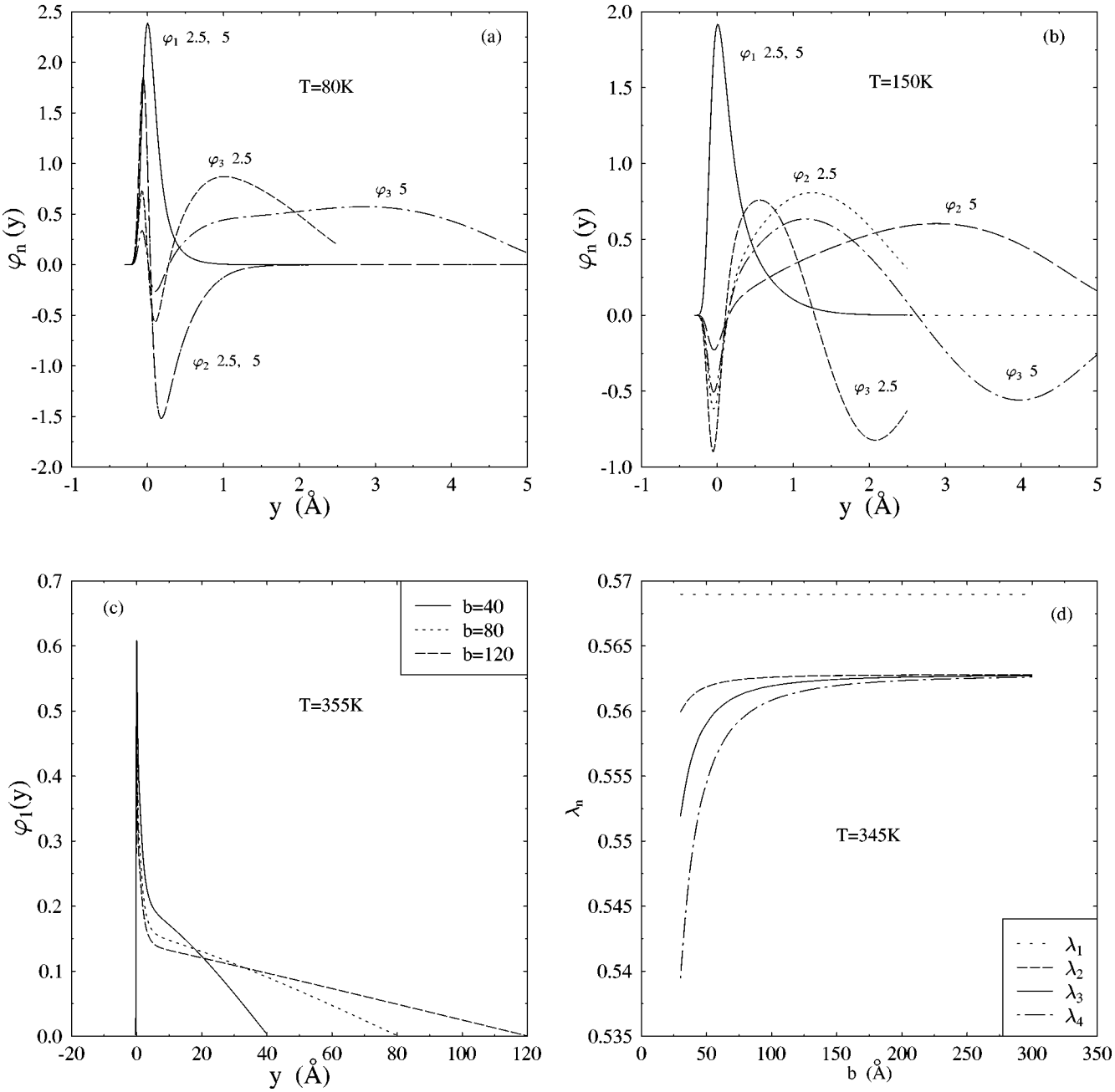

FIG. 1. (a)–(c) show the eigenfunctions calculated at different b labeled by the numbers (in Å) near the corresponding curves and different temperatures. In (a) φ_1 and φ_2 are bound states that do not change with b, while φ_3 as an unbound state does. When temperature rises to 150 K, φ_2 becomes an unbound state. (c) indicates that at $T=355$ K, which is above the melting temperature $T_m=350$ K, no bound state exists any more. (d) shows the curves of λ_n vs b. The parameters are $D=0.04$ eV, $a=4.45$ Å^{-1}, $k=0.04$ eV Å^{-2}, $\alpha=0.35$ Å^{-1}, and $\rho=0.5$.

The single-particle partition functions can be factored into contributions from external and internal degrees of freedom, viz.,

$$Z_1(V,T)=Z_{1,\mathrm{ext}}(V,T)Z_{1,\mathrm{int}}(T),$$

$$Z_2(V,T)=Z_{2,\mathrm{ext}}(V,T)Z_{2,\mathrm{int}}(T), \tag{20}$$

where $Z_{1,\mathrm{ext}}(V,T), Z_{2,\mathrm{ext}}(V,T)$ represent the translational and rotational motion of the corresponding species as a whole, whose forms were given in the Ising model of DNA melting, and $Z_{1,\mathrm{int}}(T)$ and $Z_{2,\mathrm{int}}(T)$ mainly include the vibrational degrees of freedom [1]. For the PB model,

$$Z_{1,\mathrm{int}}=\left(\frac{2\pi}{\beta k}\right)^{N/2} e^{-\beta N D} \tag{21}$$

is the partition function of a single harmonic chain moving on the plateau of the Morse potential. $Z_{2,\mathrm{int}}$ will be discussed below.

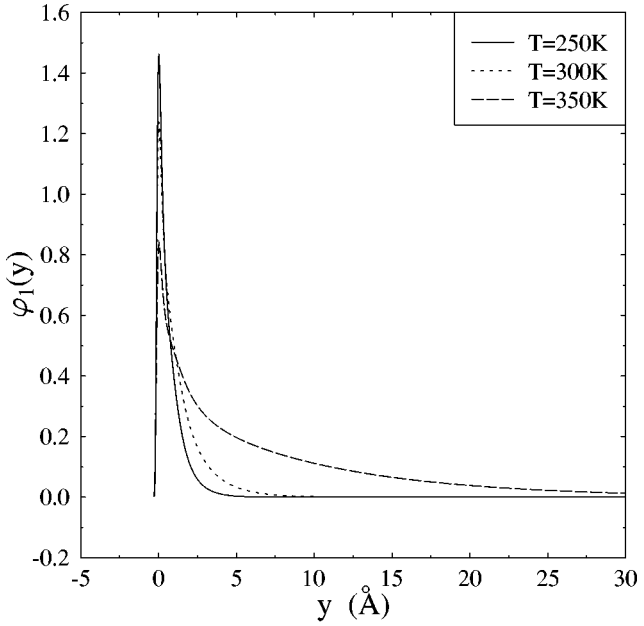

FIG. 2. The curves of φ_1 at different temperatures. Note that the sharp peak of φ_1 is related with the fact that the continuum approximation is unjustifiable for this model system.

Define θ_{ext} as the fraction of double-stranded DNA, which do not completely denature, i.e., having at least one intact base-pair. Then

$$\theta_{\text{ext}}=2N_2/N_T=\frac{4N_TK_{\text{eq}}+1-\sqrt{1+8N_TK_{\text{eq}}}}{4N_TK_{\text{eq}}}. \tag{22}$$

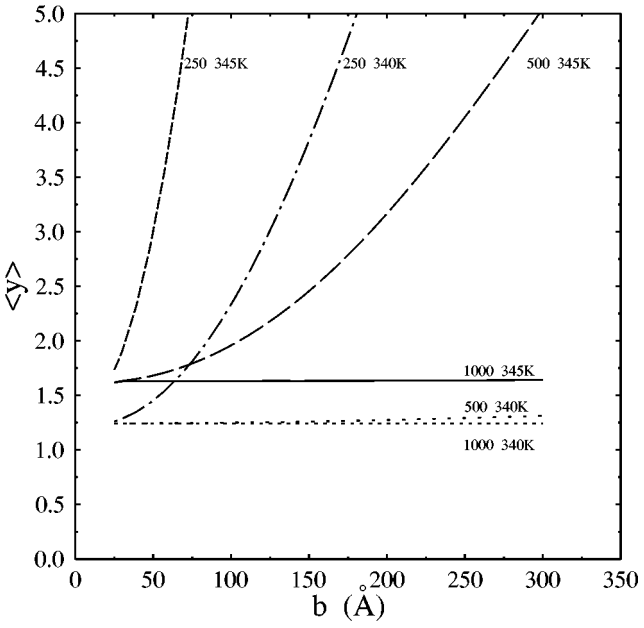

FIG. 3. The $\langle y\rangle$ vs b calculated by Eq. (44) for three values of N, 250, 500, and 1000, at two different temperatures, 340 and 345 K. Each curve is labeled by two numbers, indicating the N and the temperature, respectively. We can see that the velocity of the divergence of $\langle y\rangle$ is drastically affected by the total number of base pairs and temperature. Divergence is obvious in all cases. To get the above curves, the number of the Gauss-Legendre nodes used to solve the TI equation (7) is chosen to be 800.

E. $Z_{2,\text{int}}$, dynamic diameter and θ_{int}

Based on the above factorization of the partition functions, we find that for $Z_{2,\text{int}}$, it is only bound states that should be taken into account for the summation in Eq. (12), i.e.,

$$Z_{2,\text{int}}=\sum_{\text{bound states}}\lambda_n^N. \tag{23}$$

Correspondingly, the thermal expansion of the base pair is rewritten as

$$\langle y\rangle=\frac{1}{Z_{\text{bound states}}}\sum\langle n|y|n\rangle\lambda_n^N. \tag{24}$$

With the above modification, the divergence of the thermodynamic quantities is naturally eliminated. Note that Eqs. (23) and (24) are correct only for the homogeneous chain with periodical boundary condition where the standard TI technique is applicable. In the following, we strive to extend the formulas so that it can hold for the inhomogeneous chain with open boundaries.

From Eq. (15), we see that the thermodynamic quantity $\langle y\rangle$ exhibits an internal product form. This implies that $|\varphi_1(y)|^2$ can be regarded as the probability density whose scope represents the fluctuations of the base pairs in a DNA molecule. Figure 2 shows that this scope changes with temperature. We define the scope of the eigenvector just below the melting temperature as the *dynamic diameter* (denoted by d) of duplex DNA molecule because it is the largest separation of a base pair due to the fluctuations. When $y>d$, the probability amplitude of y is nearly zero. This suggests that, as an approximation, we can choose the subspace $[-\infty,d]$ of the phase space of the system as the range occupied by internal states, which leads to a modified PB model with Hamiltonian

$$H_y,(-\infty<y_1,y_2,\ldots,y_N<d). \tag{25}$$

The above treatment implies that the PB model only describes the internal motion of a double-stranded DNA molecule. The value of d can be estimated from Fig. 2. For the parameters $D=0.04$ eV, $a=4.45$ Å^{-1}, $k=0.04$ eV Å^{-2}, $\alpha=0.35$ Å^{-1}, and $\rho=0.5$ [17],

$$d\approx 30\text{ Å}.$$

It can be proved that, for the homogeneous chain with periodic boundaries, the above definition of the internal states in the phase space coincides with the original separation in the spectrum of the TI equation where the bound states constitute internal states. According to the former, $Z_{2,\text{int}}$ is equal to the Z in Eq. (12) and $\langle y\rangle$ is still expressed by Eq. (13), except the concerned eigenvalues and eigenvectors pertain to the TI equation with $b=d$. Because in this case all states are nondegenerate, the series in Eqs. (12) and (13) are convergent. In the numerical computation, it is found that the first ten states ensure the convergences when $N>200$. Figure 4 shows $\langle y\rangle$ for different N calculated by two kinds of methods. It can be seen that, although excited states besides $\varphi_1(y)$ are included in the summation, their contribution to $\langle y\rangle$ is negligible. If $\langle y\rangle=2$ Å is defined as a melting crite-

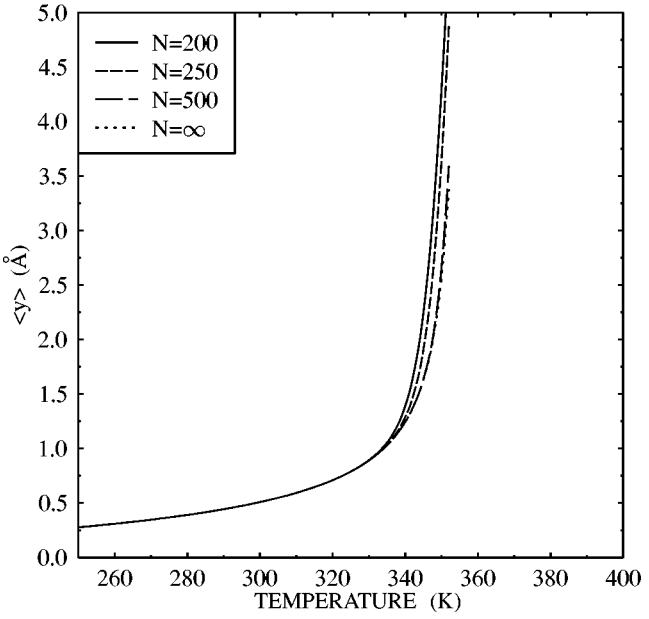

FIG. 4. $\langle y\rangle$ vs T calculated by Eq. (13) with $b=30$ Å. The curve labeled ''$N=\infty$'' is calculated by Eq. (15), which is overlapped by the curve ''$N=500$.''

rion of base pairs, the resulted shifts of the melting points are less than 3 K when $N>200$. This indicates that for long chains, the definitions of the two kinds of internal states are consistent and the introduction of d is reasonable.

Besides the chain dissociation, partly unstacking within a double-stranded DNA molecule is obviously extremely important, especially for long DNA molecules. According to the above separation of the motion of DNA molecule, this process is governed by the modified PB model. Here we give a threshold of the stacking and unstacking of base pairs in a duplex DNA molecule. When the $\langle y\rangle$ of a base pair exceeds a certain value, it can be regarded as out of stacking. For the time being, we have to determine this value from the point of the model itself. The criterion for choosing the threshold should be that when $\langle y\rangle$ exceeds the value, it diverges with slight temperature increase. We find from Fig. 4 that a threshold value of 2.0 Å is appropriate for this purpose and be used throughout the paper. The ratio of the number of the internal unstacking base pairs to the number of total base pairs is defined as $\theta_{\rm int}$.

It has been noted that, since we choose $\langle y\rangle=2$ Å, as the criterion of the melting of a base pair, the ground state of the TI equation at this critical temperature is insensitive to different parameters of the model, such as $D=0.038$ eV or $D=0.042$ eV (with other parameters unchanged). This indicates that it is reasonable for us to extend the concept of d to an inhomogeneous chain. In the following calculations, we set d to b.

In short, we consider that the melting of an ensemble of DNA molecules consists of two processes: one is the unstacking of base pairs in a single duplex DNA molecule, which is described by the modified PB model with the Hamiltonian given in formula (25) and the other is the dissociation process $C_2 \rightleftharpoons 2C_1$, which is governed by the law of mass action. The average faction of the unstacking base pairs, $\theta(T)$, of the ensemble is given by

$$\theta(T)=\theta_{\rm int}\theta_{\rm ext}. \tag{26}$$

The calculation of $\theta_{\rm ext}$ has been included in the Ising model of DNA melting. In the following we mainly consider the computation of $\theta_{\rm int}$. This corresponds to the case of long DNAs ($N>600$) because for long DNA, the internal unwinding process governs the melting transition, i.e., $\theta(T)\approx\theta_{\rm int}(T)$ [1].

III. DISCRETENESS AND NONLINEAR EFFECTS

It has been shown by direct comparisons and other calculations that the model system has strong discreteness effect and nonlinear effect [14,28], which cause significant consequences [15,29]. In this section, we investigate these effects from another point of view. The calculations are based on the self-consistent phonon approximation, taking advantage of Gaussian model [22].

In Eqs. (2) and (3), when temperature is low, the displacement y_n is small, an harmonic approximation for the Morse potential and stacking energy is legitimate. Neglecting the trivial kinetic terms, the Hamiltonian is approximately equal to

$$H_0=\sum_n\left[\frac{\phi}{2}(u_n-u_{n-1})^2+\frac{\Omega^2}{2}u_n^2\right], \tag{27}$$

where $u_n=y_n$, and ϕ and Ω^2 are defined by

$$\phi=k \tag{28}$$

and

$$\Omega^2=2a^2D. \tag{29}$$

Since the Hamiltonian H_0 has the form of a Gaussian model in statistical mechanics, its results can be conveniently used. Thus we get the free energy per base pair

$$f=-\frac{1}{2\beta}\ln\frac{8\pi}{\beta(\Omega+\sqrt{\Omega^2+4\phi})^2} \tag{30}$$

and correlation functions

$$\langle u^2\rangle\equiv\langle u_n^2\rangle=\frac{1}{\beta\Omega\sqrt{\Omega^2+4\phi}}, \tag{31}$$

$$\langle v^2\rangle\equiv\langle u_nu_{n+1}\rangle=\left(1+\frac{\Omega^2}{2\phi}\right)\langle u^2\rangle-\frac{1}{2\beta\phi}. \tag{32}$$

In order to investigate the continuum-limit approximation, we recalculate the free energy of H_0 by the TI technique. Defining the kernel

$$T(x,y)=\exp\left(-\beta\left[\frac{\phi}{2}(x-y)^2+\frac{\Omega^2}{2}x^2\right]\right) \tag{33}$$

and using the identity [30]

$$\int dy \exp\left[-\frac{1}{2t}(x-y)^2\right] f(y) = \sqrt{2\pi t}\exp\left(\frac{t}{2}\frac{\partial^2}{\partial x^2}\right) f(x) \tag{34}$$

and bearing in mind the continuum approximation, which is equal to the commutation of two operators here [23], we simplify the TI equation into an ordinary differential equation

$$\left(-\frac{1}{2\beta\phi}\frac{\partial^2}{\partial x^2}+\frac{1}{2}\beta\Omega^2 x^2\right)\varphi(x)=\left(\frac{1}{2}\ln\frac{2\pi}{\beta\phi}-\ln\lambda\right)\varphi(x). \tag{35}$$

The free energy of the system in the continuum approximation according to its largest eigenvalue is

$$f_c=-\frac{1}{2\beta}\left(\ln\frac{2\pi}{\beta\phi}-\sqrt{\frac{\Omega^2}{\phi}}\right). \tag{36}$$

Comparing Eq. (30) to Eq. (36) the condition of the continuum approximation keeping the exact free energy f close to f_c is

$$\xi\equiv\frac{\Omega^2}{2\phi}\ll 1. \tag{37}$$

This result agrees with Ref. [30].

For the typical parameters given in Sec. II E,

$$\xi\approx 13>1. \tag{38}$$

The continuum approximation does not hold here at all, which implies the system has strong discreteness effect.

In the above we assumed that the displacement y_n is small, which corresponds to the low temperature case. The analysis does not hold for higher temperature. However, taking advantage of SCPA [16,31], which is appropriate for relatively higher temperature, we can replay the treatment. In such a case, H_0 becomes an effective Hamiltonian and $u_n=y_n-\eta$ with $\eta\equiv\langle y\rangle$. Equations (28) and (29) are replaced by

$$\phi=k\{1+\rho[1-\alpha^2(\langle u^2\rangle-\langle v^2\rangle)]e^{-2\alpha\eta+\alpha^2(\langle u^2\rangle+\langle v^2\rangle)}\} \tag{39}$$

and

$$\Omega^2=2Da(2\alpha-a)e^{-a\eta+1/2\,a^2\langle u^2\rangle}+4Da(a-\alpha)e^{-2a\eta+2a^2\langle u^2\rangle}, \tag{40}$$

respectively. η is determined by equation

$$Da[e^{-a\eta+1/2\,a^2\langle u^2\rangle}-e^{-2a\eta+2a^2\langle u^2\rangle}]=\alpha k\rho[\langle u^2\rangle-\langle v^2\rangle]e^{-2\alpha\eta+\alpha^2(\langle u^2\rangle+\langle v^2\rangle)}. \tag{41}$$

Equations (31), (32), (39), (40), and (41) depend on each other and could be solved self-consistently. Figure 5 shows ξ versus temperature. It shows that for the whole temperature region calculated by SCPA, $\xi>1$, indicating the invalidity of the continuum approximation for the system. It must be noted that, because of the strong nonlinear effect, only below 200 K or so can SCPA give correct results, as shown by Fig. 6. For temperature higher than 200 K, direct comparison has been given by Dauxois *et al.* [14]. Combining the two methods, it is proved that the system exists strong discreteness effect.

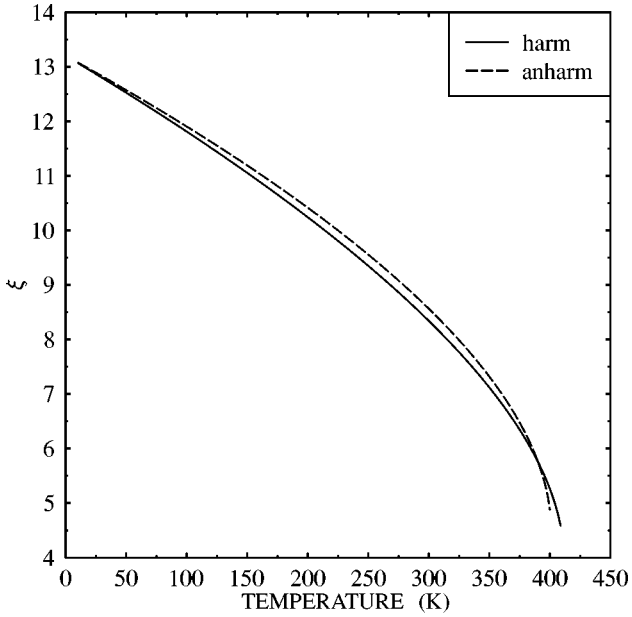

FIG. 5. ξ vs T for harmonic stacking potential ($\rho=0$) and anharmonic potential ($\rho=0.5$).

In Fig. 6, we compare $\langle y\rangle$ obtained by SCPA to the exact results obtained by the TI technique. It is shown that two sets of results are very different. For rigorous calculations, when the anharmonic stacking energy is adopted, the results are improved distinctly. But for the SCPA, the results are almost the same. This indicates the limitations of the SCPA in treating strongly nonlinear system.

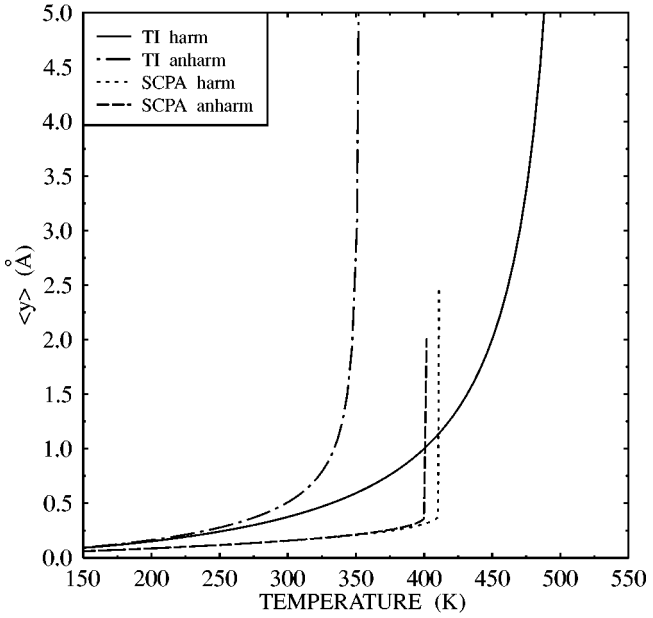

FIG. 6. The curves $\langle y\rangle$-T calculated by the rigorous TI technique or SCPA with harmonic or anharmonic stacking potential.

IV. BOUNDARY EFFECTS

So far we only consider cases with periodic boundary conditions. We do not expect different boundary conditions to be important for long chain DNAs, but for short chain ones, boundaries may have significant effects on the melting of DNAs. Because of advantages of shorter moleculers over kilo-base-pair (kbp) long molecules, such as the simplifications of the theory-experiment comparisons and convenience of controlling the systems in equilibrium states, shorter DNAs with 100−600 bp are far more often adopted experimentally. It has been pointed out that the melting process undergoes internal loops formation and end unwinding. For short DNAs, due to the high free energy costs in forming internal loops, the end meltings will govern the unwinding processes [1]. Aiming at showing boundary effects, we consider the case of open boundaries and investigate the effects of the end unwinding on the melting of DNAs with different lengths. Extension from the periodic boundary condition to open boundary condition is straightforward.

We define

$$a_m \equiv \int dy \varphi_m(y) e^{-1/2\,\beta V(y)} \tag{42}$$

and

$$y_{mn} \equiv \langle m|y|n\rangle. \tag{43}$$

As the transform invariance no longer exists, the thermal average $\langle y_i\rangle$ depends upon the positions of base pairs in the whole chain where i denotes the ith base pair. Thus

$$\langle y_i\rangle = \sum_{mn} a_m a_n y_{mn} \Delta_m^{i-1} \Delta_n^{N-i} \Big/ \sum_m a_m^2 \Delta_m^{N-1}, \tag{44}$$

where $\Delta_m \equiv \lambda_m/\lambda_1$. The above summation includes all eigenstates of the TI equation in principle. However, numerical experiment shows that inclusion of the first ten eigenstates is good enough to guarantee the convergences of $\langle y_i\rangle$.

The numerical results are presented in Fig. 7 and Fig. 8. The curves in Fig. 7 are the thermal expansion curves of two DNAs with 1000 and 3000 bp in different temperatures. It can be found that the expansions of base pairs near the boundaries are the same for long DNAs, i.e., independent of N. This fact presents the characteristics of the model with nearest-neighboring interaction. Figure 8 shows the calculated differential melting curves $d\theta_{\rm int}/dT$ for DNAs with different lengths. We can see that for short DNA, the boundary effects lead to two consequencies. In adition to widening the melting transition region, they also cause the shift of the critical point. Compared with the experimental observations, we find that the boundary effects exhibited by the model are slightly too strong. Experiments have shown that the melting region does not shift or widen so distinctly by the number of base pairs of DNA [1]. This may be a consequence of the sole nearest neighboring interaction of the base pairs or the nonoptimal parameters in the model.

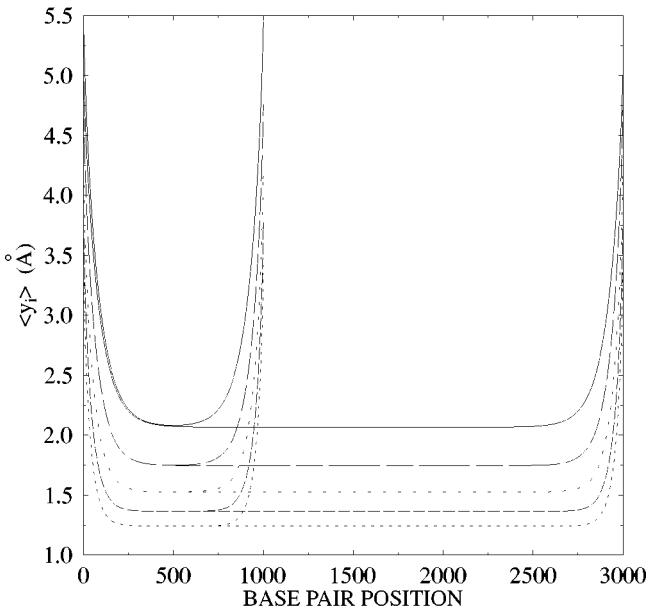

FIG. 7. $\langle y_i\rangle$ of each base pair for two homogeneous DNAs, 1000 and 3000 bp, with open boundaries at different temperatures. The curves from bottom to top correspond to $T=340$, 342, 344, 346, and 348 K, respectively.

V. EXACT SOLUTIONS FOR DNA WITH ALTERNATING BASE PAIRS

With the developments of synthetic techniques of DNAs, the melting of DNAs with alternating base pairs was carried out in experiments [32]. These samples offer the advantages of investigating the effects resulting from the differences in stacking energies. Prohofsky and co-workers have systematically investigated alternating DNAs by MSPA. In this section, we will see that with the aid of TI techniques, the PB model can be easily extended from the homogeneous chain model to periodical copolymers model without approxima-

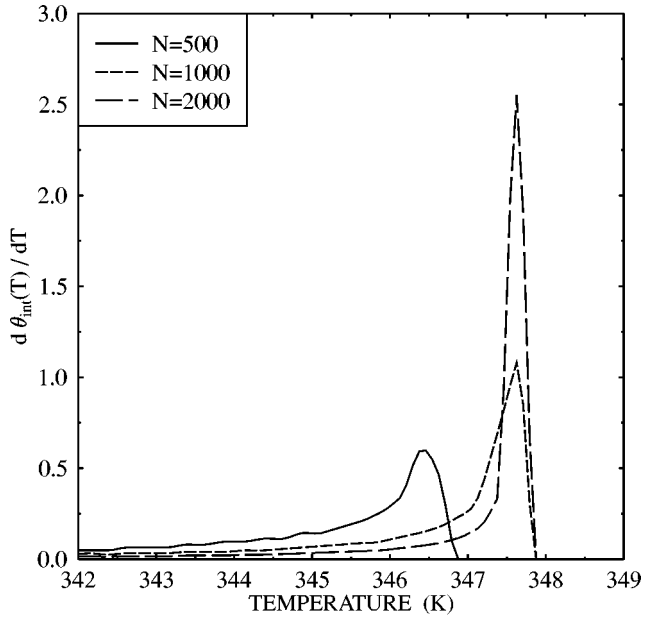

FIG. 8. Differential melting curves for homogeneous DNAs with open boundaries and different length N.

tion. Furthermore, we can take these results as criteria for our *extended transfer matrix approach* applied to DNAs with arbitrary sequences to be analyzed in next section.

Here we consider the differences of binding energies of base pairs G· C and A · T (where G, C, A, and T denote guanine, cytosine, adenine, and thymine, respectively). We use $V_s(y)$ and $V_w(y)$ to represent the hydrogen bond potentials for G· C and A· T , respectively (here *s* means ''strong'' to denote G· C base pair and *w* means ''weak'' for A· T). They are all Morse potentials, but with different parameters. For simplicity, only different D's in Eq. (2), i.e., D_s and D_w are used to take into account their differences without considering the differences of stacking energies between base pairs. In this section as well as in the next, we adopt the following parameters:

$$D_s = 0.042 \text{ eV}, \quad D_w = 0.038 \text{ eV}, \quad k = 0.042 \text{ eV/Å}$$

with other parameters unchanged.

We first consider the periodic chain with one A · T and one G· C in a unit cell. Define

$$K_{sw}(x,y) = \exp(-\beta\{w(x,y) + \tfrac{1}{2}[V_s(x) + V_w(y)]\}) \tag{45}$$

and

$$K_{ws}(x,y) = \exp(-\beta\{w(x,y) + \tfrac{1}{2}[V_w(x) + V_s(y)]\}). \tag{46}$$

It is evident that

$$K_{sw}(x,y) = K_{ws}(y,x). \tag{47}$$

Thus the partition function can be written as

$$Z = \int dy_1 dy_2 \cdots dy_N K_{sw}(y_1,y_2) \times K_{ws}(y_2,y_3) \cdots K_{ws}(y_N,y_1). \tag{48}$$

If we choose the kernel for the TI equation as

$$K(x,y) = \int dz K_{sw}(x,z) K_{ws}(z,y), \tag{49}$$

Eq. (48) can be expressed as

$$Z = \int dy_1 dy_3 \cdots K(y_1,y_3) K(y_3,y_5) \cdots K(y_{N-1},y_1). \tag{50}$$

Taking advantage of the TI technique, we have the free energy

$$f = -\frac{1}{2\beta} \ln \lambda_1. \tag{51}$$

The thermal expansions are now different for G · C and A· T base pairs,

$$\langle y_s \rangle = \int dy \varphi_1^2(y) y, \tag{52}$$

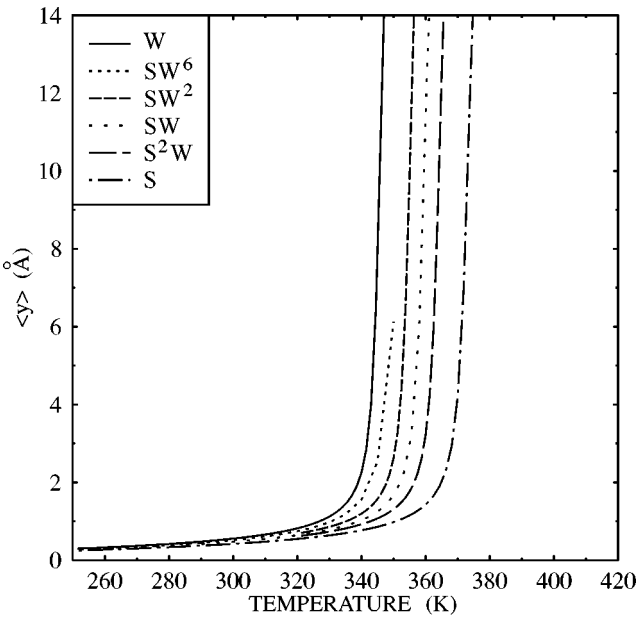

FIG. 9. $\langle y \rangle$ vs T for different periodic DNAs. The curves are denoted by the composition of base pairs in a unit cell among which the curve ''SW^6'' is calculated by the extended transfer matrix approach introduced in Sec. VI. Here ''S'' and ''W'' denote G·C and A·T, respectively. For these DNAs, the $\langle y_S \rangle \approx \langle y_W \rangle$ within relative error 5%, so they cannot be distinguished from the curves.

while $\langle y_w \rangle$ can be obtained by swapping the potentials $V_w(y)$ and $V_s(x)$. This corresponds to an alternative unit cell.

The periodic chains with unit cells s^2w, sw^2 can be handled similarly.

Figure 9 shows the calculated $\langle y \rangle$ for different chains. It implies the following results.

(i) The melting temperature is roughly proportional to the content of G · C, which is in agreement with experimental observations [33].

(ii) No subtransition exists in the melting profiles of the periodic DNAs with two or three base pairs in a unit cell, which is also in agreement with experimental observations [32].

VI. THE EXTENDED TRANSFER MATRIX APPROACH FOR ARBITRARY SEQUENCES

The most exciting phenomenon in the experiments of DNA melting are their fine structures where the melting transitions are constituted by several subtransitions. In the UV absorbances of DNAs, the fine structures are seen as a series of sharp peaks with a half-width of about 0.3–1.0 K. These facts result from the local denaturations of DNAs with inhomogeneous base-pair compositions. In biology, the phenomenon has great significance. It has been proved that the local denaturations come into being during the initiation of transcription [7]. It was also pointed out that the stability of regions within a promoter affects the efficiency of transcription. For example, the promoters tend to be embedded in relatively A · T-rich regions [34]. Furthermore, the thermal stability of local sequences is the pathway to understand the interactions between DNA molecules and biomolecules in-

cluding all kinds of drugs and proteins. Chen *et al.* have investigated the stabilities of DNA-drug complexes based on the MSPA [35]. They found that the dissociation probabilities of drugs were determined not only by the bond types connecting the drugs and DNAs, such as hydrogen bonds, covalent bonds, and Coulomb interactions, but also by the stabilities of base pairs within the regions.

In the previous section, we have shown that for the periodic DNA with two or three base pairs in a unit cell, no fine structures exist. In principle, we can consider a bigger unit cell in order to investigate the fine structures. But the numerical solutions are quite computationally demanding. In this section, we give a high precision approximate algorithm that makes it possible to calculate the melting profiles of DNAs with any sequences.

In the derivations of the partition functions by the TI technique in Sec. II, we note that Eq. (11), where the kernel is expanded by a series of orthonormal base functions, is crucial, but the form of the expansion is flexible. Based on this idea, we develop a method called the extended transfer matrix approach (ETMA), which leads to an efficient algorithm for the melting of inhomogeneous DNA and overcomes the weaknesses of molecular dynamic simulation. The latter has been widely carried out by Peyrard *et al.* [14,19,36], and Profhofsky [13] for this DNA melting model. In our method, the calculation of the thermal expansion of each base pair as well as the partition function can be reduced to the multiplicative computation of a series of matrices.

The partition function is

$$Z=\int dy_1 dy_2 \cdots dy_N K^{(1)}(y_1,y_2)K^{(2)}(y_2,y_3)\cdots K^{(N)} \times(y_N,y_1). \tag{53}$$

Here the periodic condition is not always necessary and $K^{(i)}(y_i,y_{i+1})$ is defined by

$$K^{(i)}(y_i,y_{i+1})=\exp(-\beta\{w_i(y_i,y_{i+1})+\tfrac{1}{2}[V_i(y_i)+V_{i+1}(y_{i+1})]\}), \tag{54}$$

where the stacking energy $w_i(y_i,y_{i+1})$ and the potential between a base pair $V_i(y_i)$ are site dependent. For simplicity, we have $w_i(y_i,y_{i+1})=w(y_i,y_{i+1})$ and let $V_i(y_i)$ be chosen from $V_s(y_i)$ or $V_w(y_i)$.

Two methods are given in order to carry out our ETMA, which are different in their ways of expanding the kernels $(K^{(i)}s)$.

A. Method one

In this method, different kernels are expanded by same set of bases. As noted before, $K^{(i)}(x,y)$ is defined in the space $\{a,b;a,b\}$, and its norm exists. We can choose any complete set of orthonormal base functions, $\{\varphi_m(y)\}$, to expand the $K^{(i)}(x,y)$, i.e.,

$$K^{(i)}(x,y)=\sum_{n,m=1}^{+\infty} C_{nm}^{(i)}\varphi_n(x)\varphi_m(y), \tag{55}$$

which obviously is more general than Eq. (11). The coefficient $C_{nm}^{(i)}$ is determined by

$$C_{nm}^{(i)}=\int\int dxdyK^{(i)}(x,y)\varphi_n(x)\varphi_m(y). \tag{56}$$

In practice, Eq. (55) is truncated within the first M bases. Thus Eq. (55) is rewritten as

$$K^{(i)}(x,y)\approx\sum_{n,m=1}^{M} C_{nm}^{(i)}\varphi_n(x)\varphi_m(y). \tag{57}$$

Substituting Eq. (57) in to Eq. (53), we obtain

$$Z=\mathrm{Tr}\left(\prod_{i=1}^{N}\widetilde{C}^{(i)}\right), \tag{58}$$

where the matrix $\widetilde{C}^{(i)}$ is given by

$$\widetilde{C}_{nm}^{(i)}=C_{nm}^{(i)}. \tag{59}$$

Thus the partition function becomes the products of N matrices with dimension of M. Obviously, Eq. (58) is the discreteness form of Eq. (53). Compared with the Ising model, we can see that $\widetilde{C}^{(i)}$ resembles the transfer matrix there. Due to its M dimension, Eq. (58) can be viewed as the Ising model with Ising spin having M components. Thus the PB model can be regarded as an extension of the Ising model of DNA melting. Because in the experiments on the melting of DNA, the lengths of the DNAs are rarely more than 3000 bp and M is not too large as long as the base functions are carefully chosen, Eq. (58) can be numerically solved.

If we define the matrix $\widetilde{Y}$ as

$$\widetilde{Y}_{nm}\equiv\int dy\varphi_n(y)y\varphi_m(y) \tag{60}$$

for the periodic boundary conditions, the partition function is given by Eq. (58) and the thermal expansion is

$$\langle y_i\rangle=\frac{1}{Z}\mathrm{Tr}[\widetilde{C}^{(1)}\cdots\widetilde{C}^{(i-1)}\widetilde{Y}\widetilde{C}^{(i)}\cdots\widetilde{C}^{(N)}]. \tag{61}$$

For open boundary conditions

$$Z=\mathrm{Tr}[\widetilde{C}^{(1)}\widetilde{C}^{(2)}\cdots\widetilde{C}^{(N-1)}\widetilde{A}] \tag{62}$$

and

$$\langle y_i\rangle=\frac{1}{Z}\mathrm{Tr}[\widetilde{C}^{(1)}\cdots\widetilde{C}^{(i-1)}\widetilde{Y}\widetilde{C}^{(i)}\cdots\widetilde{C}^{(N-1)}\widetilde{A}], \tag{63}$$

where the matrix $\widetilde{A}$ is defined by

$$\widetilde{A}_{nm}\equiv\int dx\varphi_n(x)e^{-1/2\,\beta V_1(x)}\int dy\varphi_m(y)e^{-1/2\,\beta V_N(y)}. \tag{64}$$

B. Method two

Unlike method one, we expand different kernels in different bases. We plan to expand the kernel to the form

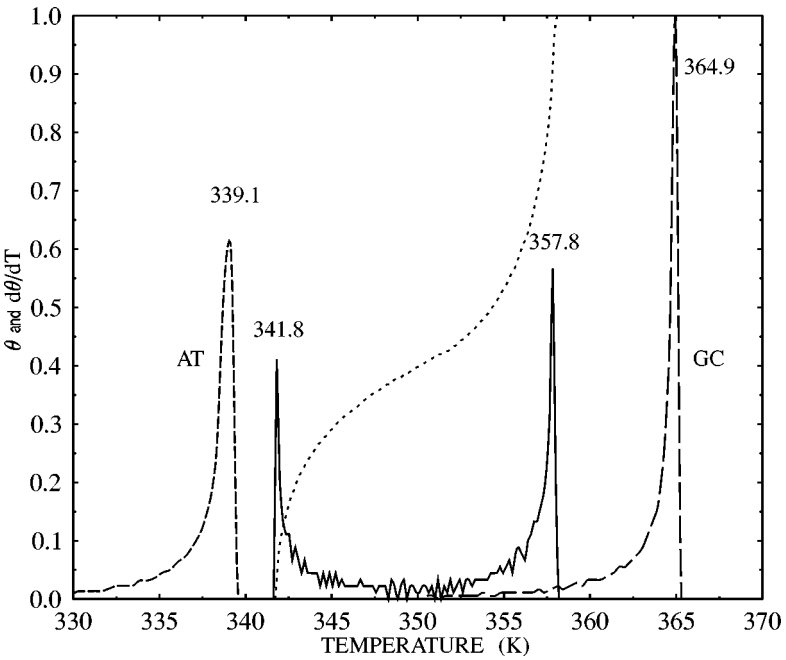

FIG. 10. Melting curve (dotted line) and differential melting curve (solid line) for 200 bp DNA. For comparison, we also show the melting curves for DNAs composed by pure A · T and G · C base pairs, which are labeled by ''AT'' and ''GC'' respectively. Each peak of the different melting curve is labeled by its corresponding temperature. The parameters are $D_a=0.042$ eV, $D_b=0.038$ eV, $a=4.45$ Å^{-1}, $k=0.042$ eV Å^{-2}, $\alpha=0.35$ Å^{-1}, and $\rho=0.5$.

$$K^{(i)}(x,y)=\sum_n \lambda_n^{(i)}\Phi_n^{(i)}(x)\Psi_n^{(i)}(y), \tag{65}$$

where $\{\Phi_n^{(i)}(x)\}$ and $\{\Psi_n^{(i)}(x)\}$ constitute complete sets of orthonormal base functions respectively, which are dependent on the form of the kernel. The approaches to expanding the four kinds of kernels $K_{ss}(x,y),K_{ww}(x,y),K_{sw}(x,y)$, $K_{ws}(x,y)$, are as follows. Due to their symmetries, the first two kernels can be directly expanded as

$$K_{ss}(x,y)=\sum_n \lambda_n^{(s)}\varphi_n^{(s)}(x)\varphi_n^{(s)}(y) \tag{66}$$

and

$$K_{ww}(x,y)=\sum_n \lambda_n^{(w)}\varphi_n^{(w)}(x)\varphi_n^{(w)}(y), \tag{67}$$

where $\{\varphi_n^{(s)}(x)\}$ and $\{\varphi_n^{(w)}(x)\}$ are given by the integral equation

$$\int dy K_{ss}(x,y)\varphi_n^{(s)}(y)=\lambda_n^{(s)}\varphi_n^{(s)}(x) \tag{68}$$

and

$$\int dy K_{ww}(x,y)\varphi_n^{(w)}(y)=\lambda_n^{(w)}\varphi_n^{(w)}(x), \tag{69}$$

respectively. For the K_{sw}, symmetrization should be applied first. We construct two symmetric kernels, $K_{sws}(x,y)$, which is the same as the $K(x,y)$ in Eq. (49) and K_{wsw} defined similarly. According to the theory of integral equation [18], there exist eigenvalues and eigenvectors satisfying

$$\int dy K_{sws}(x,y)\phi_n(y)=\lambda_n^2\phi_n(x), \tag{70}$$

$$\int dy K_{wsw}(x,y)\psi_n(y)=\lambda_n^2\psi_n(x), \tag{71}$$

and $K_{sw}(x,y)$ can be expanded as

$$K_{sw}(x,y)=\sum_n \lambda_n\phi_n(x)\psi_n(y). \tag{72}$$

Using Eq. (47), we also have the expansion of $K_{ws}(x,y)$. So in Eq. (65) $\lambda_n^{(i)}$ takes its value among $\{\lambda_n^{(s)},\lambda_n^{(w)},\lambda_n\}$, and $\Phi_n^{(i)}$ or $\Psi_n^{(i)}$ among $\{\varphi_n^{(s)},\varphi_n^{(w)},\phi_n,\psi_n\}$.

Now we define matrices $\widetilde{D}$ and $\widetilde{Z}$ whose elements are

$$\widetilde{D}_{mn}^{(i)}\equiv\sqrt{\lambda_m^{(i-1)}\lambda_n^{(i)}}\langle\Phi_m^{(i-1)}(y_i)|\Psi_n^{(i)}(y_i)\rangle \tag{73}$$

and

$$\widetilde{Z}_{mn}^{(i)}\equiv\sqrt{\lambda_m^{(i-1)}\lambda_n^{(i)}}\langle\Phi_m^{(i-1)}(y_i)|y_i|\Psi_n^{(i)}(y_i)\rangle. \tag{74}$$

respectively. According to the vectors $\Phi^{(i)}(x)$ and $\Psi^{(i)}(x)$, $\widetilde{D}^{(i)}$ and $\widetilde{Z}^{(i)}$ have eight possible forms. They are determined by $(i-1)$th, ith, and $(i+1)$th base pairs. So $\widetilde{D}^{(i)}$ chooses from $\{\widetilde{D}_{sss},\widetilde{D}_{www},\widetilde{D}_{sws},\widetilde{D}_{wsw},\widetilde{D}_{sww},\widetilde{D}_{wws},\widetilde{D}_{ssw},\widetilde{D}_{wss}\}$ where the first subscript represents the $(i-1)$th base pair, the second one the ith, and the third one the $(i+1)$th. Correspondingly, there are eight kinds of $\widetilde{Z}^{(i)}$.

For the the periodic boundary condition,

$$Z=\mathrm{Tr}[\widetilde{D}^{(1)}\cdots\widetilde{D}^{(N)}] \tag{75}$$

and

$$\langle y_i\rangle=\frac{1}{Z}\mathrm{Tr}[\widetilde{D}^{(1)}\cdots\widetilde{D}^{(i-1)}\widetilde{Z}^{(i)}\widetilde{D}^{(i+1)}\cdots\widetilde{D}^{(N)}]. \tag{76}$$

For open boundary condition,

$$Z=\mathrm{Tr}[\widetilde{B}\widetilde{D}^{(2)}\cdots\widetilde{D}^{(N-1)}] \tag{77}$$

and

$$\langle y_i\rangle = \frac{1}{Z}\,\mathrm{Tr}[\widetilde{B}\widetilde{D}^{(2)}\cdots\widetilde{D}^{(i-1)}\widetilde{Z}^{(i)}\widetilde{D}^{(i+1)}\cdots\widetilde{D}^{(N-1)}], \tag{78}$$

where $\widetilde{B}$ is defined as

$$\widetilde{B}_{mn} \equiv \sqrt{\lambda_m^{(1)}\lambda_n^{(N-1)}}\int dy_1 \Phi_m^{(1)}(y_1) e^{-1/2\,\beta V_1(y_1)} \times \int dy_N \Psi_n^{(N-1)}(y_N) e^{-1/2\,\beta V_N(y_N)}. \tag{79}$$

C. Numerical results

Carrying out the expansion (57), we find the choice of the orthonormal bases is crucial in order to have an efficient algorithm with enough precision. We have tried using a Legendre orthogonal polynomial, two-dimensional fast Fourier transformation, and the eigenvectors of the integral equation

$$\int dy K_{BB}(x,y)\varphi(y) = \lambda\varphi(x) \tag{80}$$

to expand it; the last one is found to be the best. The truncation approximation can be estimated by

$$\epsilon_i = \left| ||K^{(i)}(x,y)||^2 - \sum_{j,k=1}^{M} C_{jk}^{(i)2} \right|. \tag{81}$$

In order to avoid possible overflow in the products of the coefficient matrices, we divide every matrix that appears in both the numerator and denominator of the expression $\langle y_i\rangle$ by its largest element.

For the convenience of comparing our method with the molecular dynamics simulation in treating the melting of this DNA model, we consider the melting of block model DNAs in which A · T blocks and G · C blocks appear alternately in the chain because this is similar to the sequence pattern investigated by Dauxois *et al.* except with different parameters [19]. The two block DNAs studied below are both 1000 bp long with G · C blocks in the two boundaries, but have different block lengths: 200 bp and 40 bp.

In our calculation we first control the maximum truncation approximation, i.e.,

$$\epsilon = \max\{\epsilon_1, \epsilon_2, \ldots\}, \tag{82}$$

within the scope $\epsilon < 1.0\times10^{-6}$. This means that the minimum M is about 80 for method one and 40 for method two. With this approximation, we completely reproduce the curves shown in Fig. 7 and Fig. 9 (the relative errors are less than about 5%) , indicating that our matrix algorithm is successful and has high precision. Because the amount of computation is heavily dependent on the dimension of the matrices, the computation under the above matrix dimension is rather CPU burning. As we reduced M to detect sensitivity of the precision of $\langle y_i\rangle$, it was found that even with M as small as 30 for method one and 20 for method two, the precisions of $\langle y_i\rangle$ are intact. With these values, the CPU time for every curve shown in Figs. 12 and 13 is less than 10 min for Sparc-10 work stations. Due to the center symmetry of the base-pair sequence of the two block DNAs, only the thermal expansion of half of the base-pairs is calculated and plotted.

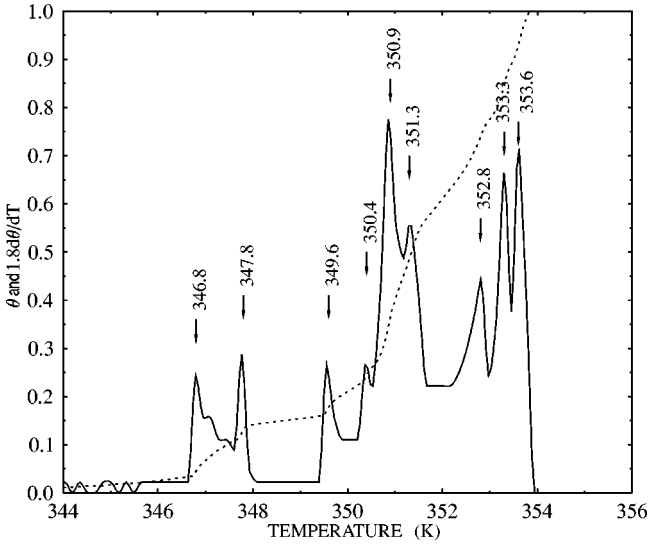

FIG. 11. Melting curve (dotted line) and differential melting curve (solid line) for 40 bp DNA. In order to show them clearly in a figure, the data for $d\theta/dT$ are magnified by a factor 1.8. The nine peaks are enumerated 1, 2, etc. from left to right.

We present the calculated melting profiles (θ and $d\theta/dT$) of the block DNAs in Figs. 10 and 11. From Fig. 10 we see clearly that for the 200 bp DNA, the melting process is constituted of two completely separate subtransitions, which correspond to the melting of A · T segments and G · C segments, respectively. This can be further verified by the thermal expansion of each base pair in different temperatures shown in Fig. 12. At the temperatures corresponding to the A · T peak and the G · C peak shown by the differential

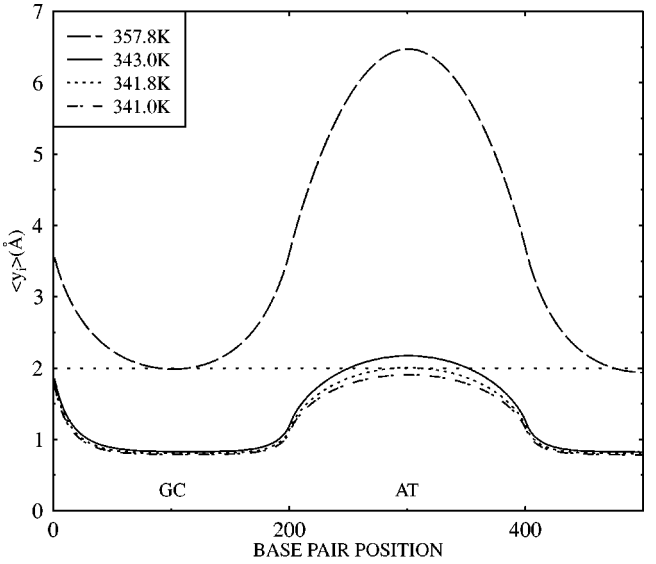

FIG. 12. The thermal expansion $\langle y_i\rangle$ of each base-pair of 200 bp DNA at the temperatures corresponding to the peaks in its differential melting curve. The dotted straight line $\langle y_i\rangle = 2$ is drawn for reference.

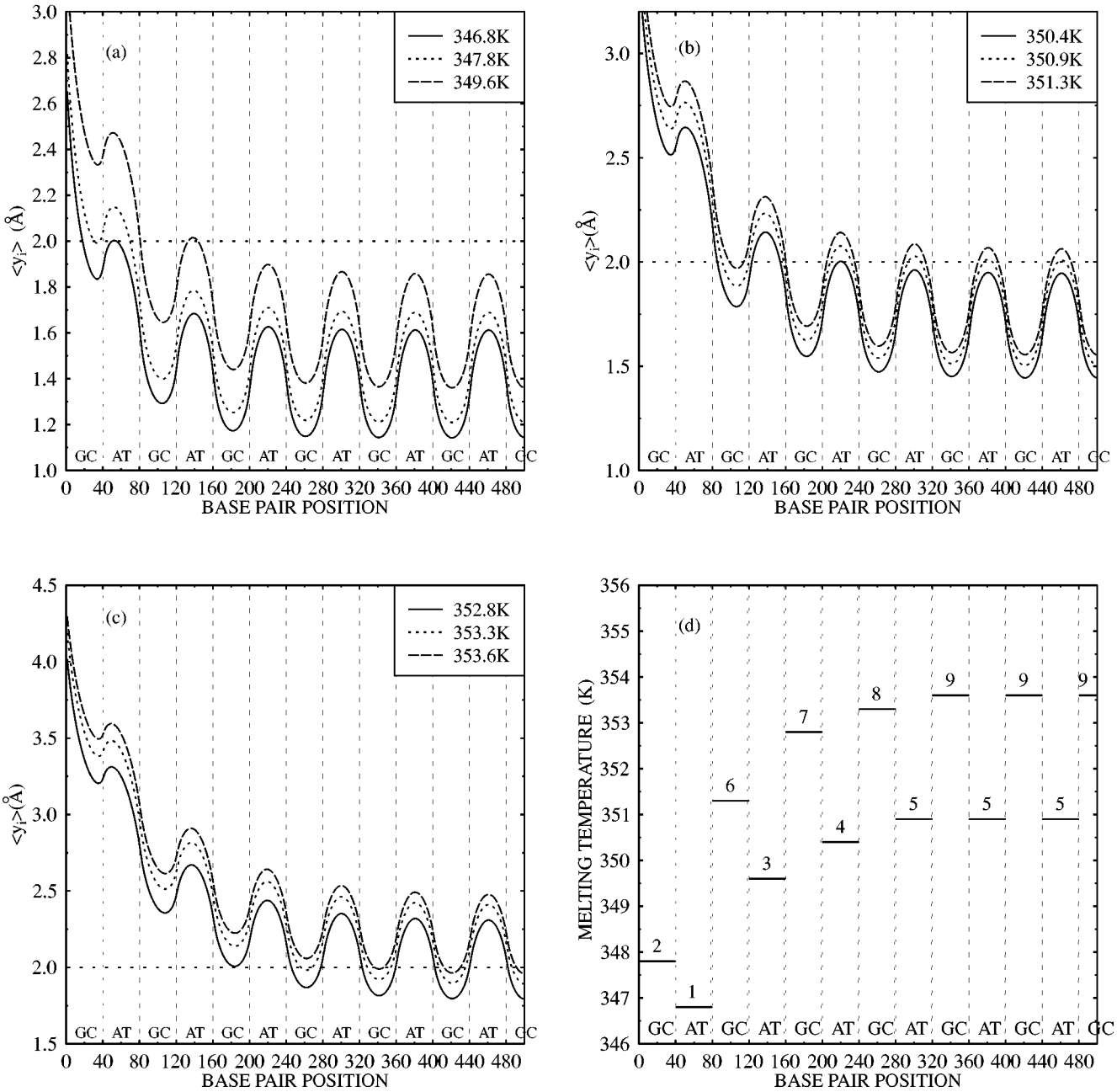

FIG. 13. (a)–(c) show the thermal expansion $\langle y_i \rangle$ of each base-pair of 40 bp DNA at the temperatures corresponding to the peaks in its differential melting curve. (d) gives the approximate melting temperatures of all segments in the 40 bp DNA and their melting sequence (labeled by the numbers near the lines).

melting curve in Fig. 10, the centers of the A · T or G · C segments just cross the line $\langle y_i \rangle = 2$ Å, which is defined as the threshold where the base pairs begin to be out of stacking. We adjust the temperature near the point corresponding to the melting of A · T segments. It was found that the thermal expansion of the base pairs in A · T segments changes distinctively while the $\langle y_i \rangle$ in G · C segments hardly shifted. This suggests that the peak is caused by the melting of A · T segments.

As can be seen from Fig. 11, the melting profile of 40 bp DNA is much more complicated than that of 200 bp DNA. Nine peaks appear in the differential melting curve. In order to find the origin of each peak, we plot the $\langle y_i \rangle$-T curve at various temperatures corresponding to every peak in Figs. 13(a)–13(c). From the points that are tangent to the line $\langle y_i \rangle = 2$, we can determine which segments produce the peak at that temperature. The melting temperature of each segment and the related peak in the differential melting curve are plotted and labeled in Fig. 13(d). From Figs. 11 and 13(d), we find that the melting of 40 bp DNA can be approximately divided into the A · T melting region (including peaks 3, 4, 5) and the G · C melting region (including peaks 7, 8, 9) like 200 bp DNA. This can be seen by estimating the areas of the corresponding peaks because they represent the ratios of the melted base pairs in the corresponding temperature regions. The shifts of peaks 1, 2, and 6 from the main

A · T or G · C region are caused by boundary effect. If we consider DNA longer than 1000 bp, we shall expect that the areas of these peaks to be less.

It must be noted that even in the same segment, different base pairs melt in rather different temperatures. Noting this feature, we find that the rigorous curve in Fig. 13(d) should be continuous. The lines only show the typical melting temperatures of the corresponding segments.

Figures 12 and 13 tell us the melting of the whole DNA begins with bubblings of the base pairs in the center regions of A · T segments, then they gradually spread to G · C segments as temperature rises. This is the very zipper mechanism of the melting of DNA presented in the Ising model of DNA melting and discussed by Peyrard and Dauxois [19] in this model. However, we must note that for the block DNA this mechanism only acts distinctly in DNAs with large blocks. With the decrease of the block sizes, the interactions between blocks become stronger, and the melting of two kinds of blocks will synthesize gradually as has been shown by the comparison between Figs. 10 and 11.

Finally, we give a comparison of our ETMA and the molecular dynamics simulation approach. From Fig. 11, we see that the resolution of ETMA can be as fine as 0.3 K, which enables it to be applied to a natural inhomogeneous DNA. To test the resolution of MD for the same model, we investigate the melting curves (θ-T) for the homogeneous DNAs given by Dauxois *et al.* [14]. For a homogeneous DNA with periodic boundaries, its rigorous melting curve should be a step function because of the translational invariance of the system. From their θ-T curve, we can see that the melting region strides over tens of Kelvin. This value can be regarded as the temperature fluctuation of the system resulted from the finite time simulating for this strongly nonlinear system. For the inhomogeneous cases, the temperature fluctuations are even larger [19]. We conclude that for the time being, the molecular dynamics simulation for this system is less competitive as far as the resolution and efficiency are considered.

VII. CONCLUSION AND DISCUSSION

In conclusion, it is noteworthy to point out the following.

(1) With the introduction of the dissociation equilibrium process between the two species of DNAs and the partition of the phase space of the system, an extended PB model is obtained.

(2) Two versions of the extended transfer matrix approaches are given, which are proved to be efficient to calculate the melting profiles of DNAs with arbitrary sequences.

(3) The PB model has a close relationship with the other two DNA melting models, i.e., the Ising model and the lattice dynamic model. According to the Ising model, it can be found that the nonlinear stacking energy term partly plays the role of loop entropy factors in the Ising model, which may show its reasonableness from a different viewpoint [1,20,30]. As compared with Prohofsky's lattice dynamic model, PB model keeps its essential nonlinear degrees of freedom though the configuration of DNA is highly simplified. This comparision suggests that if all the quadratic terms in the Hamiltonian of the lattice dynamic model were integrated out, it could be reduced to a one-dimensional model similar to the PB model. If this procedure could be executed, the empirical nonlinear stacking energy term in the PB model may get more clear physical interpretation.

(4) What we have shown in this paper is mainly on the methodology of a possible new theory of DNA melting. In order to further investigate the reasonableness of this theory, the next step should determine all parameters in the model and compare the theory with experiments.

ACKNOWLEDGMENTS

The authors thank Bai-lin Hao, Yingyao Zhou, A. R. Bishop, M. Peyrard, and C. R. Willis for kind help and valuable discussions. This work was supported in part by the National Natural Science Foundation of China and State Key Laboratory for Scientific and Engineering Computing of China.

[1] R. M. Wartell and A. S. Benight, Phys. Rep. **126**, 67 (1985).
[2] Gen-fa Zhou and Chun-Ting Zhang, Phys. Scr. **43**, 347 (1991).
[3] Y. Z. Chen and E. W. Prohofsky, Phys. Rev. E **49**, 3444 (1994).
[4] Y. Z. Chen and E. W. Prohofsky, Phys. Rev. E **51**, 5048 (1995).
[5] G. S. Manning, Q. Rev. Biophys. **II**, 179 (1978).
[6] G. Ramachandran and T. Schlick, Phys. Rev. E **51**, 6188 (1995).
[7] J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner, *Molecular Biology of the Gene* (The Benjamin/Cummings Publishing Company, Inc., Reading, MA, 1988).
[8] S. M. Lindsay, J. Powell, E. W. Prohofsky, and K. V. Devi-Prasad, in *Structure & Motion: Membranes, Nucleic Acids & Proteins*, edited by E. Clementi, G. Corongiu, M. H. Sarma, and R. H. Sarma (Adenine Press, New York, 1985).
[9] H. Grimm, H. Stiller, C. F. Majkrzak, A. Rupprecht, and U. Dahlborg, Phys. Rev. Lett. **59**, 1780 (1987).
[10] M. Peyrard and A. R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[11] P. G. Wolynes, in *Spin Glasses and Biology*, edited by D. L. Stein (Academic Press, London, 1993).
[12] D. Poland and H. A. Scheraga, *Theory of Helix-coil Transitions in Biopolymers* (Academic Press, New York, 1970).
[13] E. W. Prohofsky, *Statistical Mechanics and Stability of Macromolecules* (Cambridge University Press, Cambridge, 1995).
[14] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, 684 (1993).
[15] T. Dauxois and M. Peyrard, Phys. Rev. Lett. **70**, 3935 (1993).
[16] T. Dauxois and M. Peyrard, Phys. Rev. E **51**, 4027 (1995).
[17] T. Dauxois, M. Peyrard, and A. R. Bishop, Phys. Rev. E **47**, R44 (1993).
[18] A. C. Pipkin, *A Course on Integral Equations* (Springer, Berlin, 1991).
[19] M. Peyrard and T. Dauxois, in *Proceedings of the 6th Joint EPS-APS International Conference on Physics Computing PC '94*, edited by R. Gruber and M. Tomassini (European Physical Society, Petit-Lancy, 1994).
[20] J. R. Morris and R. J. Gooding, Phys. Rev. B **43**, 6057 (1991).

[21] J. A. Krumhansl and J. R. Schrieffer, Phys. Rev. B **11**, 3535 (1975).

[22] G. Parisi, *Statistical Field Theory* (Addison-Wesley, New York, 1988).

[23] D. J. Scalapino, M. Sears, and R. A. Ferrell, Phys. Rev. B **6**, 3409 (1972).

[24] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1992).

[25] In this case, the upper limit b is set to 30 Å. The number of points to discretize the integral will increase with the enlargement of integral range. In Dauxois and Peyrard's calculations, the b is chosen to be 190 Å. With this value of b, the dimension of the matrix will be 350×350 using Gauss-Legendre node.

[26] Landau and Lifshitz, *Statistical Physics* (Pergamon Press, Oxford, 1980).

[27] R. Peierls, *Surprises in Theoretical Physics* (Princeton University Press, Princeton, 1979).

[28] M. Techera, L. L. Daemen, and E. W. Prohofsky, Phys. Rev. A **40**, 6636 (1989).

[29] T. Dauxois, M. Peyrard, and C. R. Willis, Phys. Rev. E **48**, 4768 (1993).

[30] J. R. Morris and R. J. Gooding, J. Stat. Phys. **67**, 471 (1992).

[31] Yong-li Zhang and Wei-Mou Zheng, Phys. Rev. E **55**, 4531 (1997).

[32] R. D. Wells, J. E. Larson, R. C. Grant, B. E. Shortle, and C. R. Cantor, J. Mol. Biol. **54**, 465 (1970).

[33] W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, Berlin, 1984).

[34] S. C. Hardies, W. Hillen, T. C. Goodman, and R. D. Wells, J. Biol. Chem. **254**, 10 128 (1979).

[35] Y. Z. Chen, Yong-li Zhang, and E. W. Prohofsky, Phys. Rev. E **55**, 5843 (1997).

[36] M. Peyrard and T. Dauxois, Math. Comput. Simul. **40**, 305 (1996).