Softening of DNA near Melting as Disappearance of an Emergent Property

Debjyoti Majumdar[1, 2, *] and Somendra M. Bhattacharjee[3, †]

[1]*Institute of Physics, Bhubaneswar, Odisha, 751005, India*
[2]*Homi Bhabha National Institute, Training School Complex, Anushakti Nagar, Mumbai 400094, India*
[3]*Department of Physics, Ashoka University, Sonepat, Haryana - 131029, India*
(Dated: November 12, 2020)

Near the melting transition the bending elastic constant κ, an emergent property of double-stranded DNA (dsDNA), is shown not to follow the rodlike scaling for small length N. The reduction in κ with temperature is determined by the denatured bubbles for a continuous transition, e.g., when the two strands are Gaussian, but by the broken bonds near the open end in a Y-like configuration for a first-order transition as for strands with excluded volume interactions. In the latter case, a lever rule is operational implying a phase coexistence although dsDNA is known to be a single phase.

I. INTRODUCTION

DNA stores the genetic information in its base sequence, but its functionality relies on its physical properties, like stiffness and length. The elastic energy of DNA packaged in a viral capsid helps in the injection process[1, 2], energetically-costly bends of DNA provide sites for attachments of transcription factors and other enzymes[3–5], while the melting of DNA is a vital step in polymerase chain reactions[6]. The topological constraint when double-stranded DNA (dsDNA) is viewed as a ribbon, leads to two independent elastic constants for twist and bend[7–10]. Both these elastic constants vanish on the melting of dsDNA[7, 11, 12], when the ribbon picture is lost, showing that the stiffness is an emergent property of the bound DNA[13]. However, how this emergent behaviour goes away at melting is still unknown. Here we determine the fundamental relation between the emergent bending elastic constant κ and the fraction of broken base-pairs that drives the melting transition. We show, by simulating long semiflexible DNA, that the relation is dependent on the order of the melting transition, and involve different physical mechanisms.

For a continuous melting transition, as for Gaussian chains, a renormalized semiflexible chain picture is valid where effective κ for long chains is renormalized non-trivially by the fraction of broken bonds. Melting is found to occur homogeneously along the chain but a worm-like chain model is applicable only at low temperatures where there are no broken pairs. In contrast, in the presence of excluded volume interaction, when melting is first order, the effective κ is found to be determined by a phase-coexistence type picture with the reduction in rigidity coming mainly from the large fraction of broken bonds near the open end of dsDNA. It is not homogeneous melting, and a phase-coexistence is at odds with the conventional mechanism of bubble-induced melting transition.

The stiffness of dsDNA is expressed in terms of the persistence length[8–10] $l_p \sim 50$nm which is much larger than that of highly flexible individual strands of DNA (ssDNA) with $l_p \sim 2$nm. Generally, l_p at temperature T is defined on dimensional ground from the bending elastic constant as $l_p = \kappa/T$ (the Boltzmann constant $k_B = 1$), whereas the intuitive picture that a semiflexible polymer behaves like a rod for lengths less than l_p, follows from the decay length of the tangent-tangent correlation function, $C(i - j) = \langle \mathbf{t}_i \cdot \mathbf{t}_j \rangle \sim \exp(-|i - j|b/l_p)$, where $\mathbf{t}_i$ is the tangent to the space curve at monomer i as shown in Fig. 1, b is the bond length, and $\langle ... \rangle$ denotes ensemble average. These two definitions match for a worm-like chain which is Gaussian at long length, but not in the presence of excluded volume interaction when $C(i, j)$ decays as a power-law without any typical length[8, 14, 15].

As the base-pair energy $\sim$ 6-9 kcal/mol, thermal fluctuations lead to a cooperative breaking of the hydrogen bonds in the long length limit. This is the melting of DNA[16]. The broken base pairs may be distributed along the chain or maybe near the open end (called the Y-fork) when one end of DNA is kept fixed. A consecutive set of broken pairs is to be called a bubble; see Fig. 2. This bubble mediated transition is the usual Poland-Scheraga scheme of thermal melting of DNA[19]. The fraction n_c of unbroken bonds play the role of the order-parameter for the transition, viz., $n_c \neq 0$ ($n_c = 0$) in the dsDNA (denatured) phase, and depending on the nature of the interactions, the melting transition can be continuous or first-order[19, 20]. As the ssDNA's are flexible, the bubbles act as hinges for the rigid segments[7, 21–26], and, additionally, bubbles have biologically important roles [27, 28]. The extra flexibility introduced by the bubbles leads to a downward renormalization of the elastic constant as shown schematically in Fig. 1, provided the bubbles are distributed homogeneously along the chain. The loss of stiffness of dsDNA is gradual over a range of 20°C near melting [11], but the validity of the homogeneous picture and the functional form of the temperature dependence are not known. There is also a problem in defining $C(i, j)$ or a "ribbon" at higher temperatures for configurations dotted with bubbles (Fig. 1), unless one coarse-grains at the scale of the bubble size.

Why is dsDNA stiff when individual strands are not and how does that stiffness go away with the increase
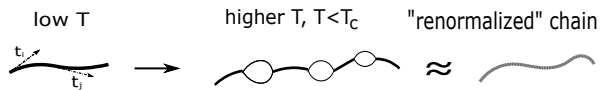
FIG. 1. Schematic diagram of renormalization of the elastic behaviour of dsDNA due to bubbles. For Gaussian chains, tangents $\mathbf{t}_i, \mathbf{t}_j$ can be used to define a persistence length (see text). Can one define a persistence length for configurations with bubbles?

FIG. 2. (a-c) Possible configurations for two successive bonds. (a) Two bound bonds with three contacts, and angle $\theta = 0$. (b) Opening of a fork. (c) same as (a) but with a bend ($\theta = \pi/2$), costing energy. There are reverse steps for Model I. E_c and E_b represent total contact and bending energies. (d) Identifying bubbles and the Y-fork.

of temperature? These questions may resemble similar ones about the rigidity of crystals. However, there are fundamental differences between the two cases. For a solid, rigidity is also an emergent phenomenon, where the shear modulus, imparting rigidity, is a consequence of continuous-symmetry breaking[13]. There is no such scenario for κ, especially because it is not the response function associated with any order-parameter, like n_c. Instead, the bound phase allows a ribbon-like description for which topological arguments[7], e.g., the Călugăreanu theorem, are applicable. The twist elastic constant, related to the helical nature, and the bending elastic constant related to the entropy of DNA[16], are relevant for dsDNA, but not for ssDNA or the denatured phase where the ribbon picture is lost. Of the two elastic constants, κ is a large scale property that should be insensitive to microscopic details, while the twist constant is dependent on the details of the structure. It is, therefore, possible to model the reduction of κ through the changes in the semiflexible bound structure mediated by the broken base-pairs. The occurrences of bubbles, as in Fig. 1, may seem to invalidate the ribbon picture, even raising questions on defining a tangent vector $\mathbf{t}$. These issues may be alleviated by coarse-graining on a scale larger than the bubble size (Fig. 1) restoring the ribbon picture with renormalized elastic constants. With this in mind, we use coarse-grained models for finite length DNA, which are important from experimental point of view, since experiments are performed upon finite system.

This paper is organized in the following manner. In Sec. II the coarse-grained models are defined. Two models, viz., Gaussian chain model and chains with self- and mutual-avoidance, both on a cubic lattice are defined there. The connection of the elastic constant with appropriate sizes via fluctuation theorems are also elaborated there. Sec. III gives the details of the simulation method of developing `PERM` for dsDNA. Sec. IV discusses the drastic difference in the behaviour of rigidity when excluded volume interactions are taken into account. Sec. V concludes the paper with some analogies between DNA melting and crystal melting.

II. MODEL AND QUALITATIVE DESCRIPTION

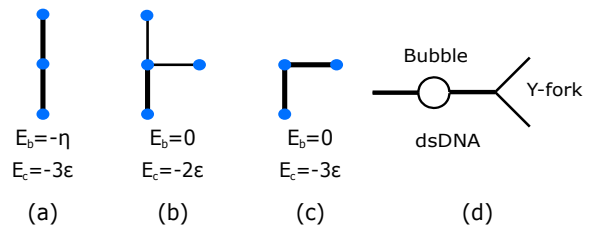Two different models are considered here, viz., I and II, on a cubic lattice. In model I, the strands are Gaussian chains, while in model II, we incorporate both self and mutual avoidance. The binding of the chains is allowed by an attractive energy $-\epsilon(\epsilon > 0)$ whenever a monomer of one chain is in contact with a monomer of the other chain provided both monomers have the same position along the chain. This ensures the native base-pairing of DNA. In each case, we consider two varieties of polymers, viz., (i) flexible polymers where both the single and double strands are flexible, and (ii) semiflexible dsDNA where only the bound parts are semiflexible but the bubbles consist of flexible chains. The semi-flexibility in dsDNA is incorporated by penalizing a bend of two successive paired bonds with energy $E_b = -\eta \cos\theta$ where θ is the angle between two bonds and $\eta(>0)$ is the bending energy constant; see Fig. 2(a)-(c). Whereas a bent ds configuration as in Fig. 2(c) is energetically favourable compared to Fig. 2(b), the latter is a source of additional entropy. Consequently, bubbles (Fig. 2(d)) are to be expected at higher temperatures vis-a-vis bent ds-chains at lower temperatures [17].

To explore the elastic behaviour, a force $\mathbf{F}$ is applied at the end point $\mathbf{r}_i(N)$ of each strand $i = 1, 2$ of length N, keeping the other ends fixed. The additional force-term in the Hamiltonian is $H_F = -\mathbf{F} \cdot \mathbf{x}$, where $\mathbf{x} = \mathbf{r}_1(N) + \mathbf{r}_2(N)$. The elastic response can be defined from a tensorial quantity χ as $\chi_{ij} = \frac{\partial \langle x_i \rangle}{\partial F_j}$, with the subscripts i, j denoting the Cartesian components. In the zero force limit ($\mathbf{F} \rightarrow 0$), isotropy can be used to define the elastic constant as $\kappa = \text{Tr}\,[\chi]$, which can be related to the zero-force fluctuations of $\mathbf{x}$ as $\bar{\kappa} \equiv k_B T \kappa = (\langle \mathbf{x}^2 \rangle - \langle \mathbf{x} \rangle^2)$, where the averaging is done with the $\mathbf{F} = 0$ Hamiltonian. This fluctuation relation allows us to determine κ without any external force. As we see, the bending elastic constant is not the response function associated with n_c and so the conventional critical behaviour of response functions in phase transition problems are not applicable here.

Naively, one may interpret $\bar{\kappa}$ as the variance of the end-to-end distance of the center-of-mass (CM) chain $\mathbf{X}(i) = [\mathbf{r}_1(i) + \mathbf{r}_2(i)]/2$. Therefore, the N dependence of $\bar{\kappa}$ is given by the size R of the CM chain, with a scaling behaviour $R \sim N^\nu$. However, the CM chain is not

expected to behave like an ordinary polymer, except in special situations, but in case it does so, the size exponent $\nu = 1/2$ for a Gaussian chain and $\nu \approx 0.588$ for a self-avoiding walk (polymers in good solvent). For a semiflexible Gaussian chain, the crossover from $R \sim N$ for $N \sim l_p$ to $R \sim \sqrt{N}$ for $N \gg l_p$, is given by[29]

$$\bar{\kappa} = 4R^2 = 8l_pN \left\{ 1 - \frac{l_p}{N} \left(1 - e^{-N/l_p} \right) \right\}, \qquad (1)$$

with l_p as the persistence length. Here both l_p and length N are measured in units of bond lengths which is set to one. This formula is used often for DNA when the CM chain more or less coincides with the strands, i.e., in absence of bubbles. In general, the temperature dependence of $\bar{\kappa}/N^{2\nu}$ would show us how DNA softens as the melting point T_c is reached.

In terms of the individual coordinates,

$$\bar{\kappa} = 2\langle r_1(N)^2 \rangle_c \left(1 + \frac{\langle \mathbf{r}_1(N) \cdot \mathbf{r}_2(N) \rangle_c}{\langle r_1(N)^2 \rangle_c} \right), \qquad (2)$$

is determined by the inter-chain correlation. In the perfect bound state with no bubbles, $\mathbf{r_1} = \mathbf{r_2}$, and we get $\bar{\kappa}/\langle r_1(N)^2 \rangle_c = 4$, while in the high-temperature phase, if the two chains remain uncorrelated, then $\bar{\kappa}$ is equal to the sum of the individual modulus. Then, $\bar{\kappa}/\langle r_1(N)^2 \rangle_c = 2$, for Gaussian chains. The ratio is expected to be > 2 for strands with excluded volume interactions because there will be inter-strand correlations for long chains as dictated by the second virial coefficient (or overlap concentration c^*) [34]. Moreover, in the bound phase, individual strands also acquire the stiffness of the state, punctuated by bubbles and Y-fork. Therefore, the microscopic stiffness would no longer be the sole parameter determining the overall elasticity of the chain (Fig. 1). In our study, it is assumed that the bending rigidity is isotropic i.e. the bending energy only depends on the angle by which the polymer is bent locally, and do not depend on the direction of bending, although, it has been shown that the bending rigidity in the direction of the grooves is essentially smaller than in the perpendicular direction[12, 33]. To characterize the transition, we computed the fraction f_b of broken bonds in the bubbles and f_Y, the fraction in the Y-fork like region. The transition temperature was determined from the specific heat curves; see Appendix A. In all cases, the order of transition is found to be independent of the value of stiffness η.

III. SIMULATION ALGORITHM

For simulation, we have used the zero parameter version of the flatPERM (Pruned and Enriched Rosenbluth Method) which generates equilibrium configurations through cloning and pruning[20, 35]. Both the strands are grown simultaneously by considering all the joint possibilities of taking steps together. The weighted atmosphere at each step, i.e., the number of free sites available for the next step, serves as the weight of that step w_n, and the weight of a configuration is the successive multiplication of the weights of the previous steps $W_N = \prod_{n=1}^{N} w_n$. For example, for the first step, each of the chains has 6 different possibilities to step into. Of a total of 36 possibilities, there are 6 possible ways of making a contact; thus the local weighted atmosphere becomes $w_1 = 30 + 6 \exp{(\epsilon/k_BT)}$. Similarly, the weight for the second step including a bend and excluded volume interaction is $w_2 = 4 \exp(\epsilon/k_BT) + \exp(\epsilon/k_BT) \exp(\eta/k_BT) + 20$. For Gaussian strands, reverse steps in the ds mode with $E_b = \eta$ are considered with an appropriate change in w_2. The partition function for chain length n is estimated by averaging over the weights of configurations of length n with respect to the number of started tours where a tour is a set of chains generated with a rooted tree topology between two successive return to the main() function. An average over 10^7 tours were used in this study for chain lengths up to 2000 and error bars are estimated on the fly; see Appendix D. Pruning and enrichment is done continuously depending on whether the ratio of the weight of the particular configuration W_n and the partition function estimate Z_n for length n, $r = W_n/Z_n$ is smaller or greater than 1 respectively. For ratio $r < 1$ the configuration keeps on growing with probability r and pruned otherwise. While for $r > 1$ we make c distinct copies with $c = min(\lfloor r \rfloor, a_n)$, where a_n is the total atmosphere $(a_n = a_1 \times a_2)$ and each copy with weight $\frac{1}{c}W_n$. And for $r = 1$ the configuration continues to grow without any pruning or enrichment. The input parameters for the simulations consist of the temperature T, contact energy ϵ, bending energy constant η with $\epsilon = k_B = 1$ throughout the simulation unless otherwise specified. To translate ΔT in our simulation to a variation in °C in experiments, one requires a proper scaling, e.g. ϵ can be estimated by comparing our T_c to the experimental melting temperature.

IV. RESULTS AND DISCUSSION

If dsDNA behaves as a semiflexible chain, then for small N, $\bar{\kappa} \sim N^2$ as a rigid rod, with a crossover to Gaussian or SAW-like behaviour for large N. No rod-like behaviour is seen for $\eta = 0$. Fig. 3(a) shows that for model I, a tightly bound DNA (without bubbles) at $\epsilon/T = 10, \eta/T = 3$ $(k_B = 1)$ satisfies Eq. (1) with $l_p = 5.2$, consistent with the estimate of l_p from a transfer matrix calculation (see Appendix F). For model II also, at low temperatures, it is possible to define a rod-like behaviour; see Fig. 3(b). However, the crossover description fails near the transition because of substantial contributions from f_b and/or f_Y. In the log-log plot, the slope for small lengths is not consistent with the rigid rod expectations. For T close to T_c, DNA is neither rodlike nor completely flexible for small chain lengths. We call this region as soft DNA. It follows that though an effec-
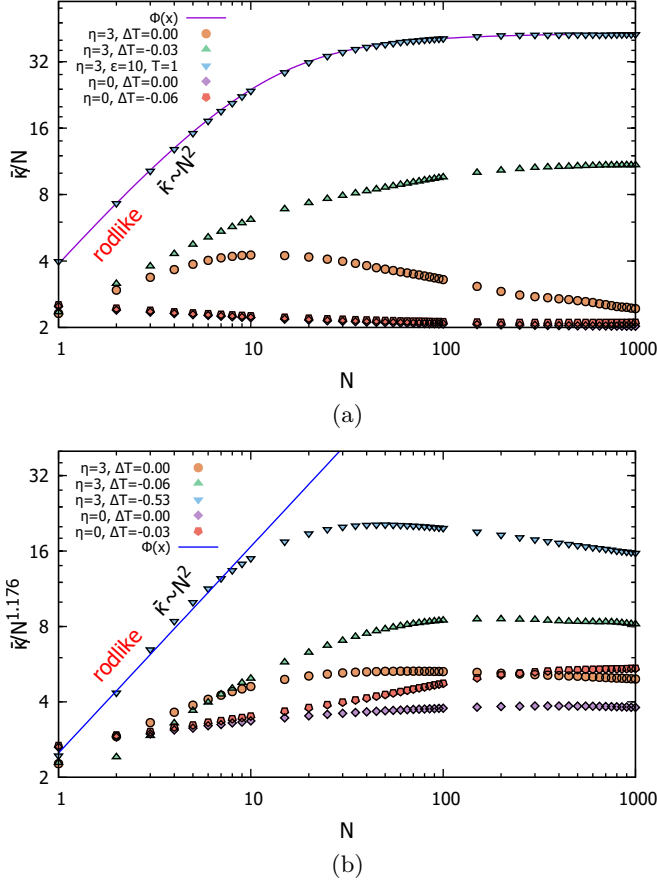
(a)

(b)

FIG. 3. (Color online) Log-log plot of $\bar{\kappa}/N^{2\nu}$ vs. N. (a) Model I for flexible $\eta = 0$ and semi-flexible chain $\eta = 3$, and different $\Delta T = T - T_c$. The curve $\phi(x)$ is a fit to the data points for $\epsilon/T = 10, \eta/T = 3$ using Eq. 1 with l_p as a parameter. (b) Model II for flexible $\eta = 0$ and semi-flexible chain $\eta = 3$, and different $\Delta T = T - T_c$. $\phi(x)$ is a straight line of slope $(2 - 2\nu)$ representing the rodlike scaling regime. No other data sets show the initial slope of $\phi(x)$.

tive elastic constant can be defined, persistence length from tangent correlations may not have any special significance.

Model I: For Gaussian chains, the melting transition is continuous at $T_c = 1.336 \pm 0.006$ for $\eta = 3$ and $T_c = 0.928 \pm 0.006$ for $\eta = 0$. Below melting, bubbles develop, and the fraction of broken bonds, $1 - n_c$, increases with temperature continuously to 1 as $T \to T_c-$ for $N \to \infty$. For finite chains, there are also broken bonds at the open end, but f_Y vs T curve sharpens into a step function for $N \to \infty$. Stiffness on the ds segments has the effect of suppressing the bubble formation at lower temperatures but the continuous transition remains intact. Fig. 4(a) shows the fractions for $\eta = 0$ and $\eta = 3$.

These results are consistent with the Poland-Scheraga picture of DNA melting, that most of the broken bonds are in the bubbles, while the fraction in the Y-region increases for $T > T_c$. As the bubbles act like hinges, the

decrease of the elastic constant with T can be attributed to the broken bonds, thereby renormalizing the effective elastic constant as in Fig. 1. As melting is continuous (see Appendix A and Fig. 4(a)) the change in the elasticity near melting is expected to be a power law $\delta\kappa \equiv \bar{\kappa}/N - (\bar{\kappa}/N)_u \sim n_c^q$ for $T \to T_c-$ [30]. The order parameter has the asymptotic behaviour $n_c \sim |(T - T_c)/T_c|^\beta$, so that the temperature dependence of elastic constant is $\delta\kappa \sim |(T - T_c)/T_c|^{q\beta}$. To extend the range of the asymptotic form valid for $T \to T_c$, we make an ansatz

$$\frac{1}{N}\bar{\kappa} = -\Delta_{\mathrm{I}}[(1 - n_c^q)^a - 1] + 2, \quad (\text{model I}) \qquad (3)$$

where Δ_{I} is the amplitude, and the exponents q and a take care of the softening by the bubbles. The values of $q = 1$ and $a = 0.1$ are found to give a good agreement of the data shown in Fig. 4(b) when the values of f_b, f_Y were used from Fig. 4(a). The same picture remains valid
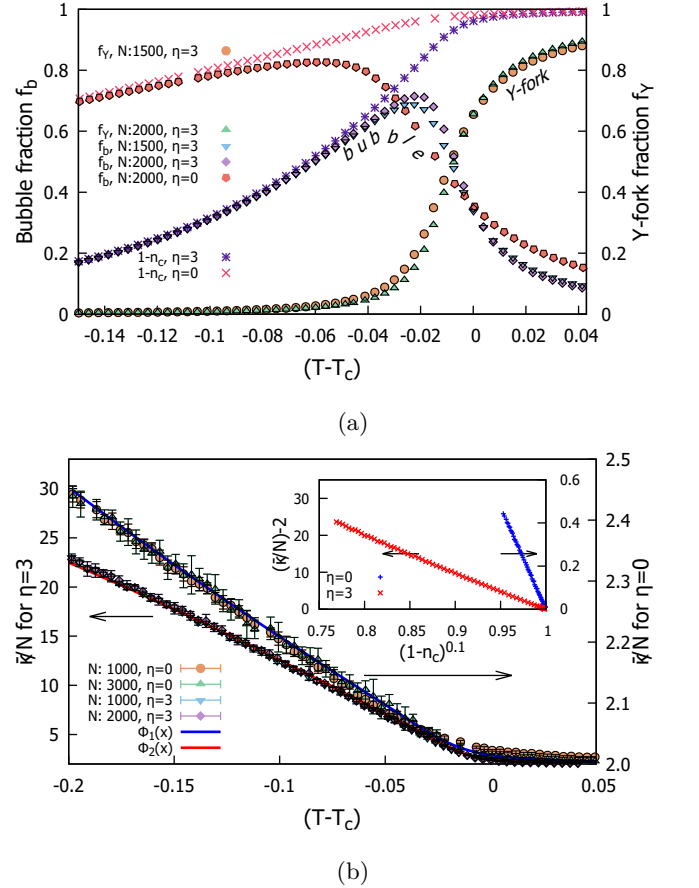

(a)

(b)

FIG. 4. (Color online) Bubble, Y and κ for $\eta = 0, 3$ for Model I (continuous transition). (a) Bubble fraction f_b, Y-fraction f_Y, and total fraction $1 - n_c$ vs $(T - T_c)$. See text for the values of T_c. (b) Rescaled elastic modulus $(\bar{\kappa}/N^{2\nu})$ vs $(T - T_c)$, with $\nu = 1/2$. Eq. (3) is shown by solid lines, ϕ_1 for $\eta = 0$ and ϕ_2 for $\eta = 3$, with n_c from (a). (Inset in (b)) Plot of $(\bar{\kappa}/N) - 2$ vs $(1 - n_c)^{0.1}$ for $\eta = 0$ and 3.
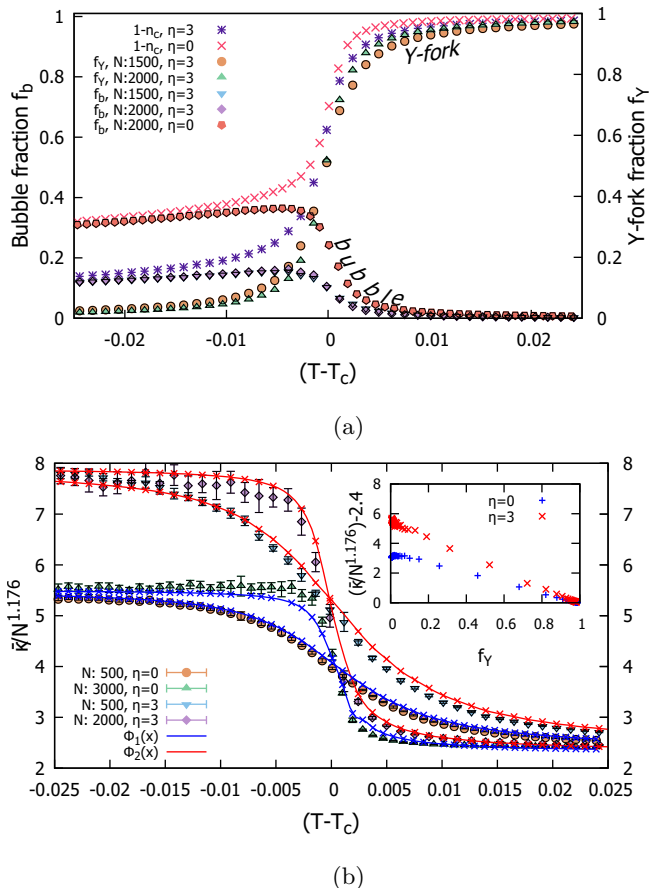
(a)

(b)

FIG. 5. (Color online) Same as in Fig. 4 but for Model II with first order transition and $\nu = 0.588$. (a) $f_b, f_Y, 1 - n_c$ vs $(T - T_c)$. In (b) $\nu = 0.588$ has been used, and Eq. 4 is shown by solid lines, ϕ_1 for $\eta = 0$ and ϕ_2 for $\eta = 3$, with f_Y taken from (a). (Inset in (b)) Plot of $(\bar{\kappa}/N^{1.176}) - 2.4$ vs f_Y for $\eta = 0$ and 3.

for both $\eta = 0$ and $\eta = 3$. We find $\Delta_{\mathrm{I}} = 9.5$ for $\eta = 0$ and $\Delta_{\mathrm{I}} = 98$ for $\eta = 3$. This proposed model for elasticity suggests that the rigidity follows the order parameter curve as the melting point is approached from the bound side.

Model II: For self and mutually-avoiding chains the melting transition is first-order at $T_c = 1.536 \pm 0.006$ for $\eta = 3$ and $T_c = 0.745 \pm 0.006$ for $\eta = 0$. The temperature dependence of f_b and f_Y are shown in Fig. 5(a), while that of $(\bar{\kappa}/N^{2\nu})$ in Fig. 5(b), where ν takes into account the effect of excluded volume interaction (see Eq. (2)). There are significant differences from model I. Close to melting, most of the broken bonds are in the Y-fork, the fraction in the bubbles remains more or less the same. Consequently, Eq. (3), encoding Fig. 1, is not meaningful; but instead an empirical equation, reminiscent of the lever rule in phase coexistence, is found to describe the data. We see a deviation from the Poland-Scheraga picture. Our proposed model for rigidity in this case is based on the superposition of the elastic constant for the bound

and the unbound Y-part in the proportion of $(1 - f_Y) : f_Y$ as

$$\bar{\kappa}/N^{2\nu} = -\Delta_{\mathrm{II}} \, f_Y^{2\nu} + (\bar{\kappa}/N^{2\nu})_{\mathrm{b}}, \quad \text{(model II)} \qquad (4)$$

which gives $(\bar{\kappa}/N^{2\nu})_{\mathrm{b}}$ of the bound phase for $f_Y \to 0$, and $(\bar{\kappa}/N^{2\nu})_{\mathrm{u}}$ of the unbound phase for $f_Y \to 1$ with $\Delta_{\mathrm{II}} = (\bar{\kappa}/N^{2\nu})_{\mathrm{b}} - (\bar{\kappa}/N^{2\nu})_{\mathrm{u}}$, the two limiting values were adjusted to get a good fit with the values of f_Y taken from Fig. 5(a). These points are also shown in the figure. The bubble contribution in the bound part is found to be very small. The temperature independence of the elastic constant on the bound side away from melting is consistent with the experimental results of [11]. The major implication of Eq. (4) is the coexistence of the bound and the unbound state, although dsDNA is a single phase.

We note that the bubble fraction f_b is lower for the semi-flexible models [Figs. 4(a) and 5(a)], and is a manifestation of the coupling between bubble formation and DNA bending energetics (see Fig. 2). For stiffer bonds, it becomes energetically favourable to maintain a bound state and make a straight move than to form a bubble to become flexible; in other words, bending energy of a semi-flexible DNA reduces the possibility of bubble formation. This tendency to maintain the bound state decreases the entropy of the system compared to the $\eta = 0$ case, thereby providing thermal stability to the bound phase. Thus the melting temperature is higher for nonzero η, and the transition becomes sharper. However, the bubble size distribution and thus the average bubble length near the transition remain unaltered by stiffness.

V. CONCLUSION

We conclude by comparing the melting picture of DNA with that of a crystal. Two main contenders of the mechanism for melting of a crystal are the homogeneous melting via the formation of defects, topological or nontopological, and a surface melting[31, 32]. If the defects form anywhere in the bulk crystal due to thermal fluctuations, the ordering is destroyed with a reduction of the rigidity. The melting process is then homogeneous. A different possibility is a surface melting where a wetting liquid layer is formed on the surface and the thickness of the layer diverges at the melting point. We do see analogs of these two processes in DNA melting, though distinctly different in detail, and dependent on the order of the transition. For continuous transitions, it is the Poland-Scheraga scheme of homogeneous bubble formation that modifies the elastic constant as in Eq. (3). For a ribbon picture to be applicable, a coarse-graining as shown in Fig. 1, is necessary. The reduction in rigidity follows the temperature dependence of the fraction of intact bonds. On the other hand, with excluded volume interaction the melting process starts at the open end, like surface melting, at temperatures below the real melting temperature, forming the Y-region. The melting process

completes when the length of the Y-region diverges (for infinite chains). In this scenario, for long chains, the density of broken bonds when expressed in terms of the length of the bound segment, viz., $f_b/(1-f_Y)$ should be independent of N, as we found to be the case (see Appendix E). This picture also suggests that if a dsDNA is capped by a sequence of high binding energy (i.e. of a higher melting point), then there is a possibility of superheating a dsDNA, when the dsDNA state can be maintained in its bound phase above the melting point. This nonequilibrium aspect is beyond the scope of equilibrium simulations.
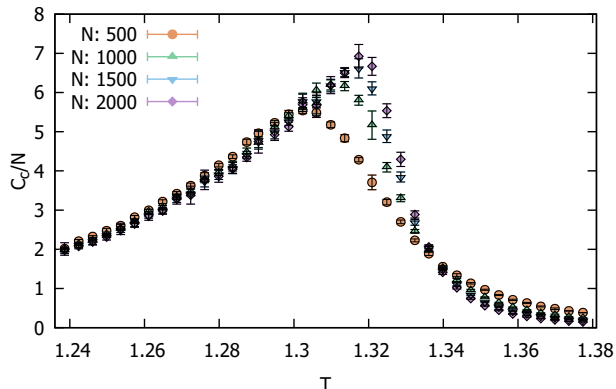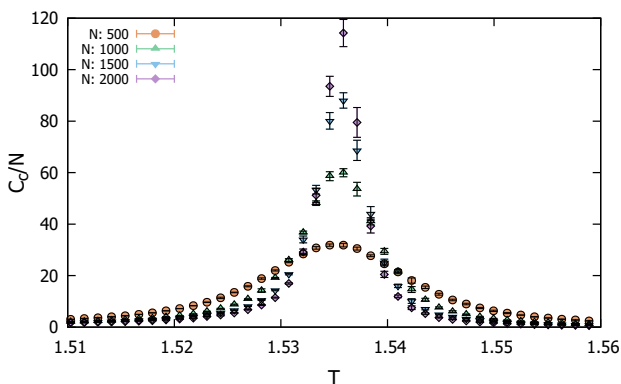
FIG. 6. (Color online) Contact number fluctuation. (a) Model I. (b) Model II. Both results are for semi-flexible models with bending energy constant $\eta = 3$.

Appendix A: Estimation of the Transition point

The transition (melting) point has been estimated from the contact number fluctuation per basepair C_c/N (related to the specific heat). For model I, undergoing a continuous transition, the transition point is determined from the intersection point of the curves for various lengths N, that remains invariant under a change of system size (Fig. 6(a)) and the transition point is estimated to be $T_c = 1.336 \pm 0.006$ for $\eta = 3$. For model II which undergoes a first-order transition, the transition point is determined from the peak of the contact number fluctuation curves (Fig. 6(b)), and the transition point is estimated to be $T_c = 1.536 \pm 0.006$ for $\eta = 3$.

Appendix B: Bubble Size Distribution

The bubble size distribution at the transition point scales as[19, 36]

$$P(l) \sim l^{-\psi} \tag{B1}$$

where $P(l)$ is the probability of a bubble of length l and the exponent ψ is related to the nature of the transition[19]. If $\psi > 2$, it represents a first-order transition, while $1 < \psi < 2$ represents a continuous transition and for $\psi < 1$ there is no transition at all. As per the definition of a bubble, broken bonds in the Y-fork is not included in the bubble statistics. See Fig. 7. The observed slopes are consistent with a continuous transition for Model I (Gaussian chains) but a first-order transition for Model II (chains with excluded volume interaction).

Appendix C: Benchmark for Simulation results

To check for the accuracy of our code in case of semi-flexible Gaussian chains, we calculated the mean squared end-to-end distance for different rigidities at a very low temperature where the DNA is in the bound state, and compared with the exact analytical result for a single ideal semi-flexible chain. Analytically, for an ideal semi-flexible chain the mean squared end-to-end distance varies with the rigidity of the local bending as[29]

$$\langle R^2 \rangle = b^2(N+1)\frac{1+L(\bar{\eta})}{1-L(\bar{\eta})} - 2b^2 L(\bar{\eta})\frac{1-L^{N+1}(\bar{\eta})}{(1-L(\bar{\eta}))^2}, \tag{C1}$$

where

$$L(\bar{\eta}) = \frac{\partial}{\partial \bar{\eta}} \ln Z(\bar{\eta}) = \frac{\sinh(\bar{\eta})}{\cosh(\bar{\eta})+2},$$

$\bar{\eta} = \eta/k_B T$ and $Z(\bar{\eta})$ is the two step partition function when no external force is applied. The comparison is shown in Fig. 8. The form of $L(\bar{\eta})$ is specific to our convention of the energetics for polymer bending, where if overlapped (i.e. in the single chain limit) the polymer
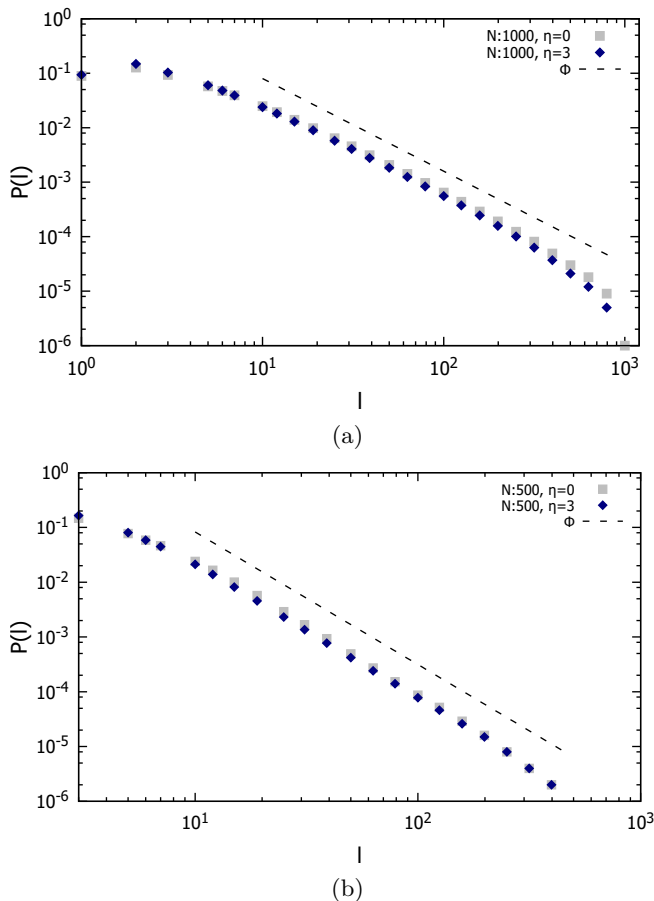
(a)

(b)

FIG. 7. (Color online) Bubble size distribution at the melting point. (a) Model I for $\eta = 0, 3$ and $T \approx T_c = .928, 1.33$ respectively, for a chain length $N = 1000$. The straight line $\phi(x) \sim x^{-1.7}$ is a fit to the intermediate region of the $\eta = 0$ data points. (b) Model II for $\eta = 0, 3$ at $T \approx T_c = .75, 1.54$ respectively, for chain length $N = 500$. The straight line $\phi(x) \sim x^{-2.4}$ is a fit to the intermediate region of the $\eta = 3$ data points.

has the following local partition function for a two step walk

$$Z = e^{\bar{\eta}} + e^{-\bar{\eta}} + 4 \tag{C2}$$

where $e^{\bar{\eta}}$ is for moving straight, $e^{-\bar{\eta}}$ for moving backwards and 1 for right angle turns over a cubic lattice according to the bending energy $E_b = -\eta \cos\theta$.

In the presence of self and mutual avoidance, for model II of DNA, a good check would be to compare the size of the polymer chain at a very low temperature (i.e. large ϵ/T) when the two chains would be completely in the overlapped state and would behave as a single rigid chain, with that of the single chain of the same rigidity[37]. The comparison is shown in Fig. 9. Note that for both dsDNA and a single self-avoiding chain, the relatve weight for bending is taken as $\exp(-\bar{\eta}) = 0.005$.
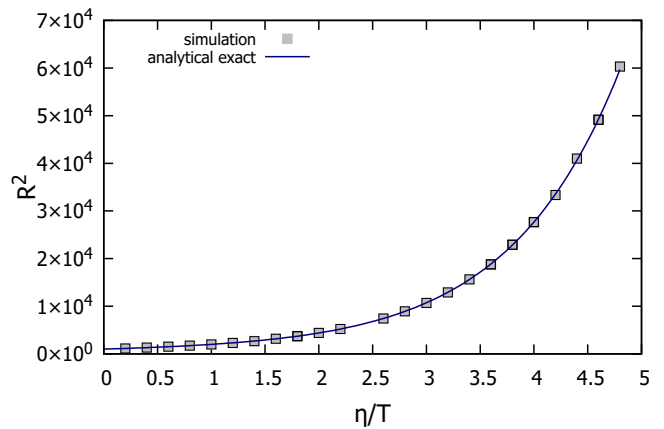

FIG. 8. (Color online) The mean squared end-to-end distance R^2 vs $\bar{\eta}$, obtained both analytically (Eq. (C1)), and from simulation at temperature T much below the $\eta = 0$ transition point T_c, $\epsilon/T = 10 \gg \epsilon/T_c = 1.077$. This is for length $N = 1000$.
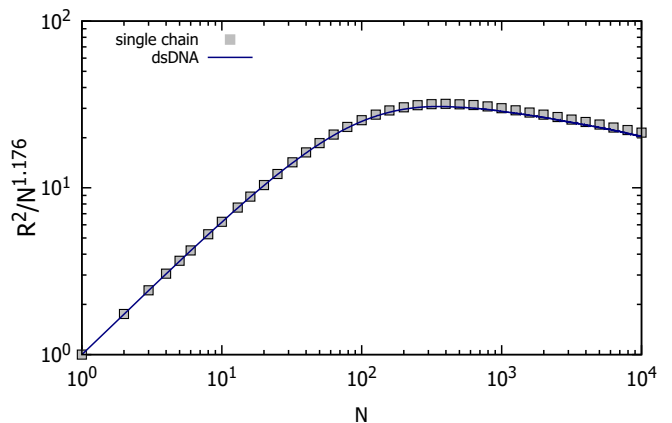

FIG. 9. (Color online) Log-log plot of the scaled mean squared end-to-end distance $R^2/N^{2\nu}$ vs. the chain length N for model II, with Boltzmann factors $\exp(\epsilon/T) = 7 \times 10^4$ for contact and $\exp(-\bar{\eta}) = 0.005$ for bends, for the dsDNA, and a single self-avoiding semi-flexible chain with the same Boltzmann factor for bending $\exp(-\bar{\eta}) = 0.005$.

Appendix D: On the fly error calculation for fluctuating quantities

The estimation of error for any thermodynamical observable is the fluctuation of that quantity. For quantities which are fluctuating in itself e.g. contact number fluctuation C_c or elastic modulus κ estimation of error becomes tricky. The way PERM is implemented every tour provides an independent estimate of any quantity which contributes to another sample in the running average. Now, with every new estimate from a tour the difference of the present estimate from the estimate of the average up to now gives a measure for the fluctuation of that quantity. The updates of the mean and the
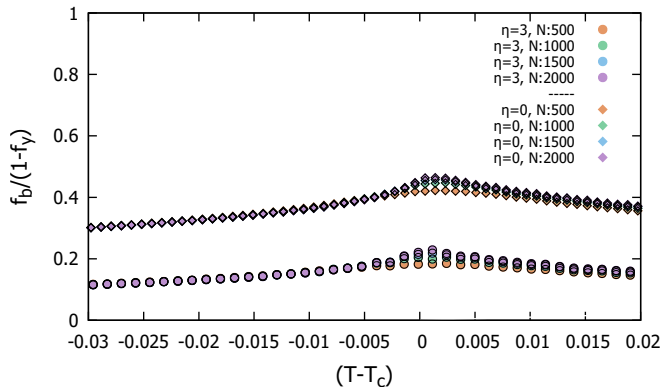
FIG. 10. (Color online) Plot of $n_b = f_b/(1 - f_Y)$ vs $T - T_c$ for $\eta = 0, 3$ and various N, using the data of Fig. 5(a). The data collapse corroborates the picture of phase coexistence. The deviations near $T = T_c$ is the usual finite size effect at a phase transition.

fluctuation of a quantity x follows the scheme

$$\overline{x}_n = \overline{x}_{n-1} + (x_n - \overline{x}_{n-1})/n \qquad (D1)$$

$$d_n^2 = d_{n-1}^2 + (x_n - \overline{x}_n)(x_n - \overline{x}_{n-1}), \qquad (D2)$$

where x_n is the current nth estimate of the quantity, $\overline{x}_{n-1}$ represents the average up to $n - 1$ samples and $d_n^2 = \sum_{i=1}^{n} (x_i - \overline{x}_n)^2$. Thus, the standard deviation is given by

$$S_n = \sqrt{\frac{d_n^2}{(n-1)}}. \qquad (D3)$$

This method is known as Welford's method.

Appendix E: Data collapse for first-order melting

Let N_b and N_Y be the total number of broken bonds in bubbles and the Y-fork region. Then $f_b = N_b/N$ and $f_Y = N_Y/N$ and their temperature dependence is shown in Fig. 5(a). If we treat the chain as consisting of two segments, bound and Y, then the density of broken bonds in the bound segment will be $n_b = N_b/(N - N_Y) = f_b/(1 - f_Y)$. For long bound segments, n_b should be independent of N. The plot of n_b vs $T - T_c$ in Fig. 10 shows a nice collapse both for $\eta = 0$ and $\eta = 3$, except for the usual finite size effect near T_c. This data collapse validates the coexistence picture.

Appendix F: Transfer Matrix Calculation of Persistence Length

For Gaussian semiflexible chains, the tangent-tangent correlation or bond-bond correlation (Fig. 1) decays exponentially for large $|i - j|$, $\langle \mathbf{t}_i \cdot \mathbf{t}_j \rangle \sim \exp(-|i - j|b/l_p)$ providing a definition for persistence length l_p, where b is the bond length. This definition is not applicable for cases with excluded volume interaction, as SAWs are critical objects [14]. The persistence length for model I with η at temperature T ($\overline{\eta} = \eta/k_B T$) can be exactly calculated from transfer matrix calculation of a two step walk. The transfer matrix for a two step walk is written as

$$\begin{pmatrix} e^{\overline{\eta}} & 1 & 1 & 1 & 1 & e^{-\overline{\eta}} \\ 1 & e^{\overline{\eta}} & 1 & 1 & e^{-\overline{\eta}} & 1 \\ 1 & 1 & e^{\overline{\eta}} & e^{-\overline{\eta}} & 1 & 1 \\ 1 & 1 & e^{-\overline{\eta}} & e^{\overline{\eta}} & 1 & 1 \\ 1 & e^{-\overline{\eta}} & 1 & 1 & e^{\overline{\eta}} & 1 \\ e^{-\overline{\eta}} & 1 & 1 & 1 & 1 & e^{\overline{\eta}} \end{pmatrix} \qquad (F1)$$

The largest eigenvalue for the above matrix is $\lambda_1 = 4 + 2\cosh\overline{\eta}$, and for $\overline{\eta} = 3$ the second largest eigenvalue obtained using MATHEMATICA is $\lambda_2 = 19.562$. Therefore, the persistence length is $l_p = [\ln(\lambda_1/\lambda_2)]^{-1} \approx 5.3$.

[1] P.K. Purohit, J. Kondev, and R. Phillips. Mechanics of DNA packaging in viruses. *Proc. Natl. Acad. Sci. USA* **100**, 3173-3178 (2003).

[2] H.G. Garcia, P. Grayson, L. Han, M. Inamdar, J. Kondev, P.C. Nelson, R. Phillips, J. Widom, and P.A. Wiggins. Biological Consequences of Tightly Bent DNA: The Other Life of a Macromolecular Celebrity. *Biopolymers* **85**, 115 (2007).

[3] M. Horikoshi, C. Bertuccioli, R. Takada, J. Wang, T. Yamamoto, and R.G. Roeder. Transcription factor TFIID induces DNA bending upon binding to the TATA element. *Proc. Natl. Acad. Sci. USA* **89**, 1060-1064 (1992).

[4] W. Rees, R. Keller, J. Vesenka, G. Yang, and C. Bustamante. Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science* **260**, 1646-1649 (1987).

[5] S. Lee, S-R. Jung, K. Heo, Jo Ann W. Byl, J.E. Deweese, N. Osheroff, and S. Hohng. DNA cleavage and opening reactions of human topoisomerase IIα are regulated via Mg²⁺-mediated dynamic bending of gate-DNA. *Proc. Natl. Acad. Sci. USA* **109**, 2925-2930 (2012).

[6] K.B. Mullis and F.A. Faloona. Specific Synthesis of DNA in Vitro via a Polymerase-Catalyzed Chain Reaction. *Methods Enzymol* **155**, 335-350 (1987).

[7] S. Geggier and A. Vologodskii. Sequence dependence of DNA bending rigidity. *Proc. Natl. Acad. Sci. USA* **107**, 15421-15426 (2010).

[8] P.J. Hagerman. Flexibility of DNA, *Ann. Rev. Biophys. Biophys. Chem.* **17**, 265 (1988).

[9] K. Rechendorff, G. Witz, J. Adamcik, and G. Dietler. Persistence length and scaling properties of single-stranded DNA adsorbed on modified graphite. *J. Chem. Phys.* **131**, 095103 (2009).

[10] J.A. Abels, F. Moreno-Herrero, T. van der Heijden, C. Dekker, and N.H. Dekker. Single-Molecule Measurements of the Persistence Length of Double-Stranded RNA. *Biophys. J.* **88**, 2737 (2005).

[11] A. Brunet, A. *et al.*. How does temperature impact the conformation of single DNA molecules below melting temperature? *Nucl. Acd. Res.* **46**, 2074 (2018).

[12] W.K. Olson, A.A. Gorin, X.J. Lu, L.M. Hock, and V.B. Zhurkin. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* **95**, 11163-11168 (1998).

[13] S. Kivelson and S.A. Kivelson. Defining emergence in physics. *npj Quantum Materials* **1**, 16024 (2016).

[14] L. Schäfer, A. Ostendorf and J. Hager. Scaling of the correlations among segment directions of a self-repelling polymer chain. *J. Phys. A: Math. Gen.* **32**, 7875 (1999).

[15] H-P. Hsu, W. Paul and K. Binder. Standard Definitions of Persistence Length Do Not Describe the Local "Intrinsic" Stiffness of Real Polymer Chains. *Macromolecules* **43**, 3094 (2010).

[16] A. Vologodskii. *Biophysics of DNA* (Cambridge University Press, Cambridge, UK, 2015).

[17] Our models differ from other models considered previously, e.g., in [18], which considered a distance dependent elastic constant. Our models differ with [18] in the following respect: (i) Both models I and II incorporate polymeric correlations in bubbles, that change the nature of the melting transition, but this correlation is not present in the model of [18]. (ii) Our Model II considers excluded volume interaction, both inter and intra strand, which are known to be relevant interactions for polymers, and change the reunion exponents relevant for melting [19]. This interaction is absent in [18].

[18] J. Palmeri, M. Manghi, and N. Destainville. Thermal Denaturation of Fluctuating DNA Driven by Bending Entropy. *Phys. Rev. Lett.* **99**, 088103 (2007).

[19] M.E. Fisher. Walks, walls, wetting, and melting. *J. Stat. Phys.* **34**, 667 (1984).

[20] M.S. Causo, B. Coluzzi, and P. Grassberger. Simple model for the DNA denaturation transition. *Phys. Rev. E* **62**, 3958 (1999).

[21] J. Yan and J.F. Marko. Localized Single-Stranded Bubble Mechanism for Cyclization of Short Double Helix DNA. *Phys. Rev. Lett.* **93**, 108108 (2004).

[22] R.A. Forties, R. Bundschuh and M.G. Poirier. The flexibility of locally melted DNA. *Nucleic Acids Res.* **37**, (2009).

[23] P.A. Wiggins, R. Phillips and P.C. Nelson. Exact theory of kinkable elastic polymers. *Phys. Rev. E* **71**, 021909 (2005).

[24] A. Vologodskii, and M.D. Frank-Kamenetskii. Strong bending of the DNA double helix. *Nucleic Acids Res.* **41**, 6785 (2013).

[25] G. Altan-Bonnet, A. Libchaber and O. Krichevsky. Bubble Dynamics in Double-Stranded DNA. *Phys. Rev. Lett.* **90**, 138101 (2003).

[26] T. Pal, T. and S.M. Bhattacharjee. Rigidity of melting DNA. *Phys. Rev. E.* **93**, 052102 (2016).

[27] C. Yuan, E. Rhoades, X.W. Lou and L.A. Archer. Spontaneous sharp bending of DNA: role of melting bubbles. *Nucleic Acids Res.* **34**, 4554 (2006).

[28] J. Ramstein and R. Lavery. Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. USA* **85**, 7231 (1988).

[29] S.M. Bhattacharjee, A. Giacometti and A. Maritan. Flory Theory for Polymers. *J. Phys. Condens. Matter* **25**, 503101 (2013).

[30] Here $(\bar{\kappa}/N)_u$ refers to the unbound phase for $N \to \infty$.

[31] S.R. Phillpot, S. Yip and D. Wolf. How Do Crystals Melt?. *Comp. in Phys.* **3**, 20 (1989).

[32] Q.S. Mei and K. Lu. Melting and superheating of crystalline solids: From bulk to nanocrystals. *Prog in Mat. Sci.* , **52** 1175-1262 (2007).

[33] V.B. Zhurkin, Y.P. Lysov and V.I. Ivanov. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.* **6**, 1081 (1979).

[34] P.G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).

[35] T. Prellberg and J. Krawczyk. Flat Histogram Version of the Pruned and Enriched Rosenbluth Method. *Phys. Rev. Lett.* **92**, 120602 (2004).

[36] E. Carlon, E. Orlandini and A.L. Stella. Roles of stiffness and excluded volume in DNA denaturation. *Phys. Rev. Lett.*, **88**, 198101 (2002).

[37] H-P. Hsu, W. Paul and K. Binder. Polymer chain stiffness vs. excluded volume: A Monte Carlo study of the crossover towards the worm-like chain model. *Euro Phys. Lett.* **92**, 28003 (2010).