Dynamic model of base pair breathing in a DNA chain with a defect

Jonathan A. D. Wattis,[1,*] Sarah A. Harris,[2,†] Christina R. Grindon,[2,‡] and Charles A. Laughton[2,§]
[1]*Division of Theoretical Mechanics, School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*
[2]*Cancer Research Laboratory, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom*
(Received 10 November 2000; published 22 May 2001)

We model and analyze a short section of a DNA chain with a defect, with the aim of understanding how the frequency, amplitude, and localization of breathing events depend on the strength of the bonds between base pairs, both along the chain and between the chains. Our results show that the presence of a defect in the chain permits the existence of a localized breather mode. The models we analyze are linear and hence solvable, with solvability extending to the statistical mechanics formulation of the problem. Parameter values for the interaction energy of a base with its nearest neighbors are obtained from AMBER. The results indicate good agreement with both the amplitude and the number of base pairs affected by defect-induced breathing motion.

DOI: 10.1103/PhysRevE.63.061903 PACS number(s): 87.14.Gg, 05.20.−y, 87.15.La, 45.50.−j

I. INTRODUCTION

The DNA duplex is characterized by the Watson-Crick pairing of opposing bases. However, one characteristic dynamical feature of DNA is the occasional local disruption of this hydrogen bonding, so that the bases become more solvent-exposed. This phenomenon is known as ''breathing.'' Taken to the extreme, breathing results in one or both of a pair of bases becoming extrahelical, which is an important mechanism by which damaged DNA bases are recognized by base excision repair (BER) enzymes. The breathing of normal base pairs takes place on the microsecond time scale and is therefore generally unobservable in atomistic simulations of DNA that are restricted to a time scale of nanoseconds due to computational cost. Recently, however, base pair breathing was detected by Cubero *et al.* [1] in the molecular-dynamics (MD) simulation of a DNA duplex containing a difluorotoluene (F) base, which is a nonpolar mimic of thymine (T). It was concluded that the reduction in the interaction energy of F with its partner, a conventional adenine (A) base, resulted in the time scale of breathing events moving into the range accessible by atomistic MD simulation. The close structural mimicry of F for T suggests that this study may yield much information on the mechanics and dynamics of DNA breathing that is relevant to the behavior of ''normal'' DNA, however there remains a need for alternative modeling methods that can address this issue directly.

Lattice models of DNA have been used by Yakushevich [2] and Lipniacki [3] to analyze the dynamics of DNA. In their simplest form, these reduce to the sine-Gordon equation, with more elaborate models including more degrees of freedom and more accurate descriptions of the potential energies. Simple lattice models are amenable to numerical simulation as well as detailed theoretical analysis, as we shall make use of here, and which has been performed previously, for example in [4]. A related approach has been taken by Cocco and Monasson [5] in the study of Raman spectroscopy measurements of DNA, taking a more detailed model, which required extensive numerical calculations to analyze.

Breathing events in DNA chains have undergone much study in recent years, most notably by Peyrard and Bishop [6,7]. Here we wish to analyze the effect of an inhomogeneity in the lattice, and whether the presence of such an inhomogeneity can account for breathing modes.

Our work is perhaps most similar to that of Salerno [8], who investigated the mobility of nonlinear waves through a DNA chain composed of a mixture of AT and CG base pairs. His model accounted for the changes in strength of the interstrand bonds, there being two bonds for an AT base pair and three for a GC base pair. Ting [9] investigated a chain with a defect in the form of an enzyme that moves down the DNA double strand. Ting's model is of the same form as ours and that of Peyrard and Bishop, namely a ladder shape with springs joining each base to its three nearest neighbors, two on the same strand and one on the opposite strand, as illustrated in Fig. 1. In this case, we create a one-dimensional model with linear springs, which proves sufficient to capture the observed behavior.

A major difficulty in the application of simple lattice models such as that proposed here is the derivation of realistic parameters describing the interaction potentials of neighboring base pairs. Only recently has this problem been tackled, most notably by Olsen *et al.* [10,11] and Chen *et al.* [12]. We overcome this problem by using the molecular-

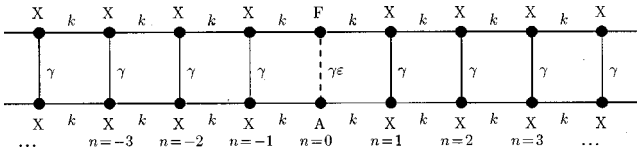

FIG. 1. Schematic of the unwound DNA double helix. *X* represents a base, horizontal lines marked *k* denote springs joining two adjacent bases and vertical lines labeled γ denote the hydrogen bonding between opposite base pairs. At the center of the chain $(n=0)$ is the inhomogeneity, a weak interchain bond denoted $\gamma\varepsilon$.

*Electronic address: Jonathan.Wattis@nottingham.ac.uk
†Electronic address: chris@holmes.cancres.nottingham.ac.uk
‡Electronic address: sarah@holmes.cancres.nottingham.ac.uk
§Electronic address: Charles.Laughton@nottingham.ac.uk

dynamics simulation package AMBER to measure the potential energy of interaction of the DNA chain in certain well-defined configurations.

In the next section, we introduce a deterministic model of the DNA double chain with a defective base, and show how it may be simplified to allow more detailed statistical analysis. We also describe how parameters required for the analysis are derived from AMBER. In Sec. III we solve the model, and instead of calculating the partition function, we find all the normal modes of oscillation. From the resulting general solution, we then find all the quantities of interest, for example rms deviation of each base pair; the more technical calculations are relegated to the Appendix. The paper concludes with a discussion of the results.

II. MODEL

We consider the DNA double helix hypothetically in its unwound state as shown in Fig. 1. We assume that each base is bonded to three neighboring bases, its base pair on the opposite chain with a weak spring of constant γ ($\gamma\varepsilon$ in the case of the inhomogeneity, with $0\leqslant\varepsilon<1$) and also bonded to its nearest neighbor on the same chain in each direction. These latter bonds are represented by stronger springs, of constant k, with $\gamma/k\ll 1$. We denote the transverse displacement of the top row of atoms by u_n and the lower by v_n, and assume that there are no longitudinal displacements. All bonds will be modeled by linear springs, yielding the Hamiltonian

$$H=\tfrac{1}{2}\gamma(\varepsilon-1)y_0^2+\sum_n \tfrac{1}{2}\dot{u}_n^2+\tfrac{1}{2}\dot{v}_n^2+\tfrac{1}{2}k(u_{n+1}-u_n)^2 + \tfrac{1}{2}k(v_{n+1}-v_n)^2+\tfrac{1}{2}\gamma(u_n-v_n)^2, \tag{2.1}$$

from which the equations of motion

$$\ddot{u}_n=k(u_{n+1}-2u_n+u_{n-1})-\gamma u_n+\gamma v_n \quad (|n|\geqslant 1), \tag{2.2}$$

$$\ddot{v}_n=k(v_{n+1}-2v_n+v_{n-1})-\gamma v_n+\gamma u_n \quad (|n|\geqslant 1), \tag{2.3}$$

$$\ddot{u}_0=k(u_1-2u_0+u_{-1})-\gamma\varepsilon u_0+\gamma\varepsilon v_0, \tag{2.4}$$

$$\ddot{v}_0=k(v_1-2v_0+v_{-1})-\gamma\varepsilon v_0+\gamma\varepsilon u_0, \tag{2.5}$$

can be derived.

We make the usual transformation of

$$x_n=u_n+v_n, \quad u_n=\tfrac{1}{2}(x_n+y_n), \qquad y_n=u_n-v_n, \quad v_n=\tfrac{1}{2}(x_n-y_n), \tag{2.6}$$

to separate out the relative motion of the interstrand distance (y_n) from the average strand displacement (x_n), the former being our main quantity of interest in this paper. This implies the equations

$$\ddot{x}_n=k(x_{n+1}-2x_n+x_{n-1}) \quad (\text{all} \quad n), \tag{2.7}$$

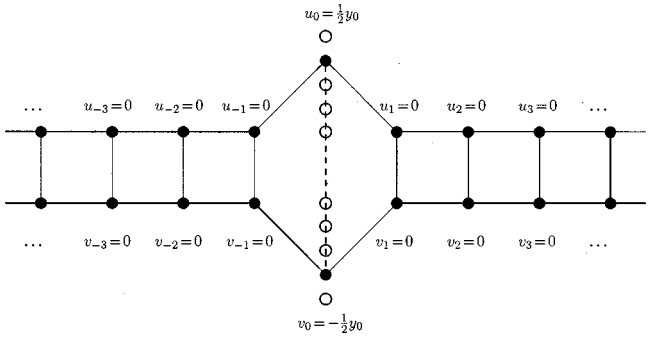

FIG. 2. Schematic of the test deformations of the DNA chain carried out in AMBER from which energy measurements are used to determine the parameters k,γ,ε. A range of y_0 values are used to build up a profile of the energy wells.

$$\ddot{y}_n=k(y_{n+1}-2y_n+y_{n-1})-2\gamma y_n \quad (|n|\geqslant 1), \tag{2.8}$$

$$\ddot{y}_0=k[\,y_1-2y_0+y_{-1}\,]-2\gamma\varepsilon y_0, \tag{2.9}$$

which demonstrate the decoupling (into x_n and y_n) of the previously coupled equations of motion (2.2)–(2.5). In these variables, the Hamiltonian (2.1) separates into $H=H_x(\dot{x}_n,x_n)+H_y(\dot{y}_n,y_n)$, where

$$H_x=\sum_n \{\tfrac{1}{2}\dot{x}_n^2+\tfrac{1}{2}k(x_{n+1}-x_n)^2\}, \quad H_y=\gamma(\varepsilon-1)y_0^2 + \sum_n \{\tfrac{1}{2}\dot{y}_n^2+\tfrac{1}{2}k(y_{n+1}-y_n)^2+\gamma y_n^2\}. \tag{2.10}$$

A. Determination of parameters

The parameters k, γ, and ε are found through the AMBER forcefield, using a normal sequence of 12 DNA base pairs and a defective sequence in which a thymine base is replaced with a difluorotoluene base. In each case, all base pairs are positioned at their minimum energy configurations, and the central base pair (AT in the normal sequence and AF in the defective chain) is then opened a certain amount, as illustrated in Fig. 2. At each choice of opening, the length of the H bond is measured, and the energy due to the AF or AT interaction is measured and the energy due to along-chain interactions is also measured. In our model, this is equivalent to positioning a DNA molecule in the configuration $y_n=0$ for all $n\neq 0$ and $y_n=y_*$ for various y_*. Using Eq. (2.6), we find $u_0=\frac{1}{2}y_*$ and $v_0=-\frac{1}{2}y_*$. In this configuration, the energy due to the cross-chain interactions is $E_\gamma=\frac{1}{2}\gamma y_*^2$, whereas the energy due to along-chain interactions is $E_k=\frac{1}{2}ky_*^2$.

The curves generated for E_k are displayed in Fig. 3, along with quadratic fits. In chains with AT and AF central base pairs, a fit of the form $A(Y-Y_0)^2+C$ has been sought, where Y_0 is the known distance between base pairs at minimum energy. The parameter A then corresponds to $\frac{1}{2}k$, and yields $k=0.0376$ J m^{-2} (using the conversion 1 Kcal M^{-1}Å^{-2}=0.6955$\times 10^{-3}$ J m^{-2}). At larger base pair openings, the backbone interactions cause the energy curve

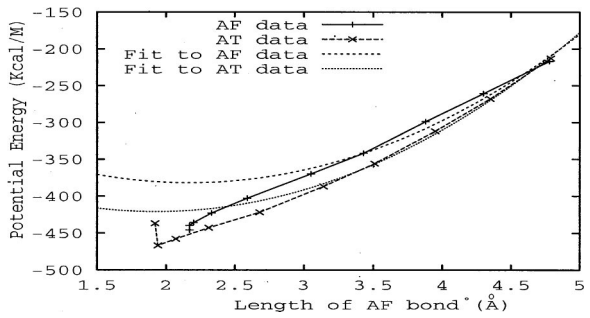

FIG. 3. Plots of the potential energy of interaction due to extension of the k springs, showing data calculated from the AMBER forcefield and quadratic fits. Along the horizontal axis, distance is measured in angstroms, and energy is measured in KcalM^{-1} up the vertical axis.

to become concave rather than convex. For this reason, rather than trying to fit a quadratic to a small portion of the data, we simply use the maximum and minimum energies, and fit a quadratic of the form $A(Y-Y_0)^2+C$ to these two points. This leads to $\gamma=0.0146\ \mathrm{J\,m}^{-2}$ and $\varepsilon=0.5$, since the values for E_γ differ considerably (recall that the data for E_k for the chain with the defective AF base pair were very similar to those for the chain containing the normal AT pair).

The mass of a base is taken as $307/N_A$ g, where N_A is Avogadro's number, giving $m=0.5098\times10^{-24}$ kg. Since all bases are assumed to have the same mass, we shall rescale time so that the mass can be ignored in the subsequent calculations. This corresponds to a rescaling of time by $\sqrt{m}=0.7\times10^{-12}$, so that each of our time units corresponds to 0.7 ps.

III. SOLUTION OF THE MODEL

In this section, we find all the normal modes of this system of equations, and determine their amplitudes by setting the energy in each mode to be $\frac{1}{2}k_BT$. We consider a finite chain whose atoms are labeled from $n=-N$ to $n=N-1$. Thus we have $2N$ base pairs, and a total of $4N$ bases. The energy calculations are simple to evaluate in the large-N limit, so this is the case on which we shall concentrate. However, there is another large quantity in our calculations, namely k/γ, since the bonds between the strands (γ) are much weaker than those along the strands (k). Thus we shall consider the regime $k/\gamma\gg N\gg1$.

To pose a fully and properly determined problem, we shall impose periodic boundary conditions, treating $x_N\equiv x_{-N}$ and $y_N\equiv y_{-N}$. Since the motion of each base is governed by a second-order ordinary differential equation, there should be $8N$ normal modes of the system: $4N$ modes in the x_n variables and $4N$ in the y_n variables. We shall find all of them. The modes appear in pairs, with a $\cos(\omega t)$ and a $\sin(\omega t)$ formally being counted as two modes even though they will share the same spatial form. Most modes are delocalized, or spatially extended, but a few are localized with larger amplitude at the inhomogeneity. We term these ''breather'' modes.

A. Normal modes for x_n

Since the x_n equations are completely independent of the inhomogeneity, we can seek the dispersion relation by assuming $x_n=Ce^{in\theta+i\omega t}$; then Eq. (2.7) implies

$$\omega^{(x)}(\theta)=2\sqrt{k}\sin(\tfrac{1}{2}\theta). \tag{3.1}$$

These are all delocalized, or spatially extended, linear waves, whose energy is spread evenly throughout the lattice. On imposing periodic boundary conditions $x_{-N}(t)\equiv x_N(t)$, we find $e^{iN\theta}=e^{-iN\theta+2im\pi}$ for some $m\in\mathbb{Z}$. Thus allowable wave numbers θ are ''quantized'' and can only take the values $\theta_m=m\pi/N$ for $m=0,1,2,\ldots,N$. The modes are then

$$x^{(\mathrm{odd})}_{n,(m)}(t)=C^{\mathrm{odd}}_m\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}(\omega^{(x)}_m t)\sin(nm\pi/N),$$

$$m=1,2,\ldots,N-1, \tag{3.2}$$

$$x^{(\mathrm{even})}_{n,(m)}(t)=C^{\mathrm{even}}_m\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}(\omega^{(x)}_m t)\cos(nm\pi/N),$$

$$m=0,1,2,\ldots,N, \tag{3.3}$$

where

$$\omega^{(x)}_m=\omega^{(x)}(\theta_m)=2\sqrt{k}\sin(m\pi/2N),\quad m=0,1,2,\ldots,N. \tag{3.4}$$

This constitutes the full complement of $4N$ modes for the second-order system of $2N$ variables $x_n(t)$ for $n=-N,\ldots,N-1$. Odd modes corresponding to $m=0$ and $m=N$ are ignored since they have zero displacement and therefore correspond to no motion. In the even case $[\cos(mn\pi/N)]$, the $m=0$ mode corresponds to motion of the whole chain with no relative motion of atoms present; this mode has zero frequency, which is due to the whole system not having any absolute reference point. The even $m=N$ mode corresponds to the highest-frequency mode where adjacent atoms move exactly out of phase.

The amplitudes $C^{\mathrm{odd}}_m, C^{\mathrm{even}}_m$ are determined by requiring the energy in each mode to be $\frac{1}{2}k_BT$. Equation (2.10) yields $\frac{1}{2}k_BT=H_x=2NC^2k\sin^2(m\pi/2N)$. Thus we have

$$C^{\mathrm{odd}}_m=C^{\mathrm{even}}_m=\frac{\sqrt{k_BT}}{2\sqrt{kN}\sin(m\pi/2N)}, \tag{3.5}$$

and the full solutions are

$$x^{(\mathrm{odd})}_{n,(m)}(t)=\frac{\sqrt{k_BT}}{2\sqrt{kN}\sin(m\pi/2N)}\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}(\omega^{(x)}_m t)\ \ \sin\left(\frac{nm\pi}{N}\right),$$

$$m=1,2,\ldots,N-1, \tag{3.6}$$

$$x^{(\mathrm{even})}_{n,(m)}(t)=\frac{\sqrt{k_BT}}{2\sqrt{kN}\sin(m\pi/2N)}\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}(\omega^{(x)}_m t)\ \ \cos\left(\frac{nm\pi}{N}\right),$$

$$m=0,1,2,\ldots,N, \tag{3.7}$$

where m enumerates the mode number. Adding the contributions from the $\sin(\omega_m^{(x)}t)$ and the $\cos(\omega_m^{(x)}t)$ modes, these solutions have mean-square displacement

$$\langle x_{n,(m)}^{(\mathrm{odd})\ 2}\rangle_t=\frac{k_BT}{8kN\sin^2(m\pi/2N)}\ \sin^2\left(\frac{nm\pi}{N}\right),$$

$$m=1,2,\ldots,N-1, \tag{3.8}$$

$$\langle x_{n,(m)}^{(\mathrm{even})\ 2}\rangle_t=\frac{k_BT}{8kN\sin^2(m\pi/2N)}\ \cos^2\left(\frac{nm\pi}{N}\right),$$

$$m=0,1,2,\ldots,N, \tag{3.9}$$

[where we use the notation $\langle f(t)\rangle_t$ for the time-averaged value of $f(t)$]. The latter formula (3.9) produces a divergent total mean-square displacement when summed over all modes (m) in the large-N limit, as can be seen by the calculation

$$\langle x_n^{(\mathrm{even})\ 2}\rangle_t=\frac{k_BT}{8kN}\sum_{m=1}^{N-1}\frac{\cos^2(nm\pi/N)}{\sin^2(m\pi/2N)}$$

$$\approx\frac{k_BT}{4k\pi}\int_0^{\pi/2}\frac{\cos^2(2nu)}{\sin^2(u)}\ du, \tag{3.10}$$

in which the singularity in the integrand at $u=0$ causes the integral to diverge. This is a standard result: there is no long-range order in one dimension at any $T>0$. The sin modes give

$$\langle x_n^{(\mathrm{odd})\ 2}\rangle_t=\frac{k_BT}{8kN}\sum_{m=1}^{N-1}\frac{\sin^2(nm\pi/N)}{\sin^2(m\pi/2N)}$$

$$\approx\frac{k_BT}{4k\pi}\int_0^{\pi/2}\frac{\sin^2(2nu)}{\sin^2(u)}\ du=\frac{k_BTn}{4k}. \tag{3.11}$$

B. Odd normal modes for y_n

We now look for the normal modes for the $y_n(t)$ variables, in which the inhomogeneity is present; this includes a breather solution. In the presence of an inhomogeneity, finding a full set of normal modes is nontrivial, so to simplify the problem, we split it into two parts noting the invariance of the determining equations under the map $n\mapsto -n$. Thus we shall look for modes that share this symmetry, namely ''even'' modes that satisfy $y_{-n}(t)=y_n(t)$, and ''odd'' modes that are antisymmetric $y_{-n}(t)=-y_n(t)$. We seek the latter first.

The condition $y_{-n}=-y_n$ immediately implies $y_0=0$ so the inhomogeneity is not noticed (as occurs for all the x modes). Inserting $y_n=Ce^{i\omega t}e^{-\lambda|n|}\sin(n\theta)$ into Eqs. (2.7) and (2.8), we find that the former is automatically satisfied, and the latter implies either $\lambda=0$ or $\theta=0,\pi$. Both $\theta=0$ and $\theta=\pi$ imply $y_n(t)\equiv 0$ and so are irrelevant; $\lambda=0$ yields the dispersion relation

$$\omega^{(y)\ 2}(\theta)=2\gamma+4k\sin^2(\tfrac{1}{2}\theta). \tag{3.12}$$

Imposing periodic boundary conditions again yields a ''quantization'' condition on θ: $y_N(t)\equiv y_{-N}(t)$, which implies $\sin(N\theta)=0$ so $\theta=\theta_m=m\pi/N$ for $m=0,1,\ldots,N$, and there are only a finite number of normal modes. Thus we have solutions

$$y_{n,(m)}^{(\mathrm{odd})}=C\sin\left(\frac{nm\pi}{N}\right)\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}(\omega_m^{(y)}t),\quad m=1,2,\ldots,N-1, \tag{3.13}$$

where $\omega_m^{(y)}=\omega^{(y)}(\theta_m)$, and C is determined by forcing the total energy of the mode to be $\frac{1}{2}k_BT$. Here we reject the modes corresponding to $m=0,\pi$ since they both imply zero displacement for the $y_n(t)$'s therefore they are irrelevant.

Equation (2.10) implies $\frac{1}{2}k_BT=H=\frac{1}{2}C^2N\omega_m^{(y)\ 2}$. Thus the mean-square displacement at site n due to the pair of modes numbered m [one for $\sin(\omega_m^{(y)}t)$ and one for $\cos(\omega_m^{(y)}t)$] is

$$\langle y_{n,(m)}^{(\mathrm{odd})\ 2}\rangle_t=\frac{k_BT\sin^2(mn\pi/N)}{4kN\left(\dfrac{\gamma}{2k}+\sin^2(m\pi/2N)\right)}. \tag{3.14}$$

Summing Eq. (3.14) over m, we can find the mean-square deviation at a particular site n due to the sum of all odd y modes (m) (See Appendix A 1 for details),

$$\langle y_n^{(\mathrm{odd})\ 2}\rangle=\frac{k_BT}{4kN}\sum_{m=1}^{N-1}\frac{\sin^2(mn\pi/N)}{(\gamma/2k)+\sin^2(m\pi/2N)}$$

$$\approx\frac{k_BT}{4k}\int_{\mu=0}^{1}\frac{1-\cos(2n\mu\pi)\ \ d\mu}{\dfrac{\gamma}{k}+1-\cos(\mu\pi)}$$

$$\approx\frac{k_BT}{4\sqrt{2\gamma k}}\left[1-\exp\left(-2\left|n\right|\sqrt{\frac{2\gamma}{k}}\right)\right], \tag{3.15}$$

where $\mu=mn/N$. Thus the odd modes have small mean-square displacement near the inhomogeneity.

C. Even normal modes for y_n

In this case, the symmetry we impose is that of $y_n(t)\equiv y_{-n}(t)$, so that the shape of the mode is invariant under $n\mapsto -n$. The inhomogeneity now plays a crucial role in our analysis. Let us assume the modes have the form

$$y_n^{(\mathrm{even})}(t)=Cf_{|n|}\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}(\omega t) \tag{3.16}$$

for some shape of oscillation determined by f_n with $f_0=1$. The displacement of the atoms nearest to the inhomogeneity ($y_{\pm 1}$) is then determined by f_1 and Eq. (2.9) implies

$$f_1=1+\frac{\gamma\varepsilon}{k}-\frac{\omega^2}{2k}. \tag{3.17}$$

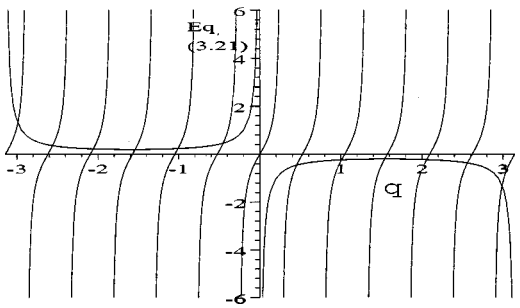

FIG. 4. Both sides of Eq. (3.21) plotted against θ showing how periodic boundary conditions quantize the allowable values for θ. In the example illustrated, $N=6$, thus our chain has 12 atoms, and there are 12 intersections of the curves, illustrating that there are as many even modes as atoms.

The displacement of all subsequent atoms is then determined by Eq. (2.8), which implies

$$0=f_{n+1}-\left(2+\frac{2\gamma}{k}-\frac{\omega^2}{k}\right)f_n+f_{n-1} \tag{3.18}$$

for $n>1$. This recurrence relation is solved by

$$f_n=C_1e^{in\theta}+C_2e^{-in\theta}, \tag{3.19}$$

where θ is related to ω through the same dispersion relation (3.12) as for odd y modes. Imposing the conditions $f_0=1$ and Eq. (3.17) on Eq. (3.19), we find the solution

$$f_n=\cos n\theta+\frac{\sin n\theta}{\sin\theta}\left(1+\frac{\gamma\varepsilon}{k}-\frac{\omega^2}{2k}-\cos\theta\right). \tag{3.20}$$

We have found a one-parameter family of solutions, parametrized by θ. This is reduced to a finite set upon imposing periodic boundary conditions at $n=\pm N$. Since we have explicitly constructed the mode to be even, the condition $y_N(t)\equiv y_{-N}(t)$ is automatically satisfied, hence to impose periodic boundary conditions we require that the joining of $n=N$ to $n=-N$ is smooth; that is, we impose $y_{N+1}(t)\equiv y_{1-N}(t)$, which implies $f_{N+1}=f_{N-1}$. This simplifies to

$$\tan(N\theta)=\frac{-\gamma(1-\varepsilon)}{k\sin\theta}. \tag{3.21}$$

Both sides of this expression are plotted in Fig. 4 using $N=6$ as an example.

Fixing N and taking the limit of small $\gamma\varepsilon/k$ (later we will find that this corresponds to large N and small γ/k with $N\ll k/\gamma$) leads to all solutions of Eq. (3.21) occurring at small values of $\tan(N\theta)$. This implies $\gamma N/k\ll 1$, and then the allowable modes are determined by

$$\theta_m=\frac{m\pi}{N}\left(1-\frac{\gamma(1-\varepsilon)}{m\pi k\sin(m\pi/N)}\right),\quad m=1,2,\ldots,N-1, \tag{3.22}$$

$$\theta_N=\pi-\sqrt{\frac{\gamma(1-\varepsilon)}{k}}. \tag{3.23}$$

The case $\tan(N\theta)=O(1)$ can only be solved numerically, if needed; here we choose to pursue theoretical results, so we shall continue with the case in which $\theta_m\sim m\pi/N$.

Combining Eqs. (3.12), (3.20), and (3.16) yields

$$y_{n,(m)}^{(\mathrm{even})}=C\left[\cos(n\theta_m)-\frac{\gamma(1-\varepsilon)\sin(|n|\theta_m)}{k\sin(\theta_m)}\right]\sin(\omega_m^{(y)}t). \tag{3.24}$$

Since in this case $\gamma N/k\ll 1$, the $\sin(|n|\theta_m)$ term is asymptotically small for all $m=1,2,\ldots,N-1$, and setting the energy of each mode to $\frac{1}{2}k_BT$ yields $C^2=k_BT/N\omega_m^{(y)\,2}$. Thus

$$y_{n,(m)}^{(\mathrm{even})}=\sqrt{\frac{k_BT}{N\omega_m^{(y)\,2}}}\left[\cos(n\theta_m)-\frac{\gamma(1-\varepsilon)\sin(|n|\theta_m)}{k\sin(\theta_m)}\right]\left\{\begin{matrix}\sin\\ \cos\end{matrix}\right\}(\omega_m^{(y)}t), \tag{3.25}$$

and so, combining the $\sin(\omega_m^{(y)}t)$ and $\cos(\omega_m^{(y)}t)$ modes, we find

$$\langle y_{n,(m)}^{(\mathrm{even})\,2}\rangle_t\sim\frac{k_BT}{2N\omega_m^{(y)\,2}}\left(1+\cos(2n\theta_m)-\frac{4\gamma(1-\varepsilon)\sin(2|n|\theta_m)}{k\sin(\theta_m)}\right). \tag{3.26}$$

To find the total mean-square displacement at site n due to all even modes, we now need to sum this over $m=1,2,\ldots,N-1$, giving

$$\begin{aligned}\langle y_n^{(\mathrm{even})\,2}\rangle&=\frac{k_BT}{4kN}\sum_{m=1}^{N-1}\frac{1+\cos(2n\theta_m)}{(\gamma/k)+1-\cos(\theta_m)}-\frac{4\gamma(1-\varepsilon)\sin(2|n|\theta_m)}{k\sin(\theta_m)\,[\,(\gamma/k)+1-\cos(\theta_m)\,]}\\&\approx\frac{k_BT}{4k}\int_{\mu=0}^{1}\frac{1+\cos(2n\mu\pi)}{(\gamma/k)+1-\cos(\mu\pi)}-\frac{4\gamma(1-\varepsilon)\sin(2|n|\mu\pi)}{k\sin(\mu\pi)\,[\,(\gamma/k)+1-\cos(\mu\pi)\,]}\,d\mu\\&\sim\frac{k_BT}{4\sqrt{2k\gamma}}\left((1+e^{-2n\sqrt{2\gamma/k}})-\sqrt{\frac{2\gamma}{k}}(1-\varepsilon)(1-e^{-2n\sqrt{2\gamma/k}})\right),\end{aligned} \tag{3.27}$$

using results from Appendix A and noting that $\langle y_n^{(\mathrm{even})\,2}\rangle=(k_BT/8k)[j_0^{(r)}+j_{2n}^{(r)}-4\gamma(1-\varepsilon)\,\eta_{2n}^{(i)}/k]$, where $q=\gamma/2k$.

D. Breathers in the infinite lattice

In long chains, it is natural to expect a mode that is centered on the defect and has an amplitude that decays with distance from the defect. Such a mode is termed ''localized,'' and since it will have a similar form to the breather

mode in the sine-Gordon equation, we use the term ''breather mode'' here, even though our system is linear.

For the simplest form of breather, we assume that the decay is exponential and monotonic, that is,

$$y_n = Ce^{-\lambda|n|}\begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(\omega t). \tag{3.28}$$

Substituting this into Eq. (2.8) leads to the dispersion relation

$$\omega_{\rm ac}^2 = 2\gamma + 2k(1-\cosh\lambda), \tag{3.29}$$

where the subscript ''ac'' stands for acoustic, since this mode bifurcates from the minimum ($\theta=0$) of the dispersion relation $\omega^{(y)}(\theta)$ (3.12). The value of λ is determined by requiring Eq.(2.9) to hold. This yields

$$\frac{\gamma(1-\varepsilon)}{k} = \sinh\lambda \sim \lambda + \tfrac{1}{6}\lambda^3 \tag{3.30}$$

for small λ; thus $\lambda \sim \gamma(1-\varepsilon)/k$, which we have already stated is a small quantity. Setting the energy of this breather mode to $\frac{1}{2}k_BT$ enables C to be determined. Thus we have the solution

$$y_n = \sqrt{\frac{k_BT(1-\varepsilon)}{2k}}\,e^{-\gamma(1-\varepsilon)|n|/k}\begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(\sqrt{2\gamma}\,t), \tag{3.31}$$

which together have a mean-square displacement at site n of

$$\langle y_n^2\rangle = \frac{k_BT(1-\varepsilon)}{2k}\,e^{-2\gamma(1-\varepsilon)|n|/k}. \tag{3.32}$$

This solution (3.31) is relevant to our study, but does not satisfy periodic boundary conditions. In the next section, we show how to modify the above analysis to produce a mode that satisfies the periodic boundary conditions and the equations of motion.

An alternative form of breather is possible, in which alternate base pairs are displaced in opposite directions. This can be described by

$$y_n = C(-1)^n e^{-\lambda|n|}\begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(\omega t). \tag{3.33}$$

As with the acoustic mode, the relationship between λ and ω is found from Eq. (2.8),

$$\omega_{\rm opt}^2 = 2\gamma + 2k(1+\cosh\lambda). \tag{3.34}$$

At $\lambda=0$, this equals the maximum frequency of the spatially extended modes (3.12), which occurs at $\theta=\pi$, thus this mode is optical in nature. Equation (2.9) determines the relevant λ, via $\sinh\lambda = -\gamma(1-\varepsilon)/k$, thus this mode is irrelevant in our case where $\varepsilon<1$. Modes that are optical in $n>0$ and acoustic in $n<0$ are impossible due to the dispersion relations (3.29) and (3.34) not overlapping.

E. Breathers in the finite lattice

The observant reader will note that in Sec. (III C) we found $2N$ even modes and $2N-2$ odd modes in Sec. (III B), leaving a deficit of two modes (since there should be $4N$ modes in a second-order system with $2N$ degrees of freedom). These final modes come from a breather-type oscillation as discussed in Sec. III D that is, an even localized mode centered on the inhomogeneity. However, the mode found there (3.31) does not quite satisfy the periodic boundary conditions. The discrepancy is small, but we now show that a similar mode that satisfies the boundary conditions can be constructed. Note that in Sec. III C we have restricted ourselves to theoretical analysis of the case $N \ll k/\gamma$, thus the solution derived below where this condition is kept will not necessarily agree with the $N\to\infty$ case considered in the preceding section.

We seek a time-harmonic solution of Eq. (2.8), as in Eq. (3.16), but assume exponential growth and decay in n, which leads to a solution of the form $y_n = A\cosh(\lambda n) + B\sinh(\lambda n)$. Here, ω is related to λ through Eq. (3.29) as in the previous calculation of a breather mode. Imposing $f_0=1$ and Eq. (3.17) implies

$$f_n = \cosh(\lambda n) - \frac{\gamma(1-\varepsilon)}{k\sinh(\lambda)}\sinh(\lambda n) = \frac{\cosh(\lambda(N-|n|))}{\cosh(\lambda N)}. \tag{3.35}$$

The value of λ is determined by imposing periodic boundary conditions and, assuming $N \ll k/\gamma$, this gives $\lambda^2 = \gamma(1-\varepsilon)/kN$. The constant C is determined by requiring the energy in the mode to be $\frac{1}{2}k_BT$ giving $C^2 = k_BT/2\gamma N$, thus

$$y_n^{(\rm br)} = \sqrt{\frac{k_BT}{4\gamma N}}\cosh\left(\sqrt{\frac{2\gamma(1-\varepsilon)}{kN}}(N-|n|)\right)\begin{Bmatrix}\sin\\ \cos\end{Bmatrix}(\sqrt{2\gamma}t) \tag{3.36}$$

and thus

$$\langle y_n^{(\rm br)\ 2}\rangle_t = \frac{k_BT}{2\gamma N}\cosh^2\left(\sqrt{\frac{2\gamma(1-\varepsilon)}{kN}}(N-|n|)\right). \tag{3.37}$$

In order for this oscillation to be observable, we need to verify that the extended/nonlocal modes have average amplitudes that sum to a smaller total. Thus, in the next section we shall compare the mean-square deviations for all modes.

F. Summary

In the large chain length (N) limit, the average displacement of the atoms ($x=u+v$) diverges (3.10), as is to be expected for a system with continuous degrees of freedom. The separation between the two chains ($y=u-v$), however, is well-behaved. We calculated this in several parts: the odd modes (3.15), the extended even modes (3.27), and the breather mode (3.32). Now we put all these together to work

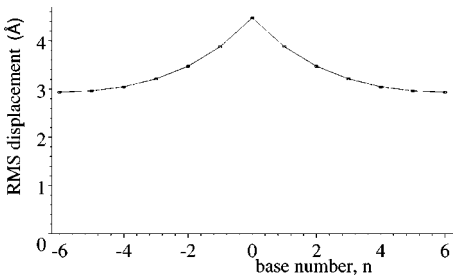

FIG. 5. A plot of the rms displacement for the distance between base pairs (3.38) measured in Å , against base number for the case $\varepsilon=0.5$, $\gamma=0.0146\ \mathrm{J\,m^{-2}}$, $k=0.0376\ \mathrm{J\,m^{-2}}$, $N=6$, and $T=300$ K.

out the total mean-square displacement due to all modes, and find how this depends on the distance from the defect (n),

$$\langle y_n^2\rangle=\langle y_n^{(\mathrm{odd})\ 2}\rangle+\langle y_n^{(\mathrm{even})\ 2}\rangle+\langle y_n^{(\mathrm{br})\ 2}\rangle$$
$$=\frac{k_BT}{4\sqrt{2k\gamma}}(1-e^{-2|n|\sqrt{2\gamma/k}})+\frac{k_BT}{4\sqrt{2k\gamma}}$$
$$\times\left[1+e^{-2|n|\sqrt{2\gamma/k}}+O\left(\frac{\gamma}{k}\right)\right]+\frac{k_BT}{2\gamma N}\cosh^2(\lambda(N-|n|))$$
$$=\frac{k_BT}{2\sqrt{2k\gamma}}\left[1+\sqrt{\frac{2k}{N^2\gamma}}\right.$$
$$\left.\times\cosh^2\left(\sqrt{\frac{2\gamma(1-\varepsilon)}{kN}}(N-|n|)\right)\right]. \qquad (3.38)$$

This quantity is plotted in Fig. 5, for parameters derived from the AMBER energies derived earlier (Sec. II A). In this case, we see that the central base (where the defect is) deviates from its equilibrium position about one and a half times as far as the bases away from the defect. This increased deviation only affects the atoms immediately neighboring the defect ($n=\pm1$), and becomes negligible at the second-neighbor sites ($n=\pm2$) . In exceedingly long chains, away from the defect, the rms deviation of the interchain spacing from equilibrium is $\sqrt{k_BT}/\sqrt[4]{(8k\gamma)}$. Near the defect, there is an increase in this displacement. Counterintuitively, this increase in displacement, although caused by $\varepsilon\neq1$, is not very sensitive to the value of ε, but rather on $\sqrt{2k/N^2\gamma}$. Depending on the exact system being studied, this could be highly significant, or very small, since our analysis rests on $1/N$ being small and $k/N\gamma$ being large, and this amplitude is the product of these two quantities. The number of bases affected by this is $O(\sqrt{(kN/\gamma(1-\varepsilon))})$.

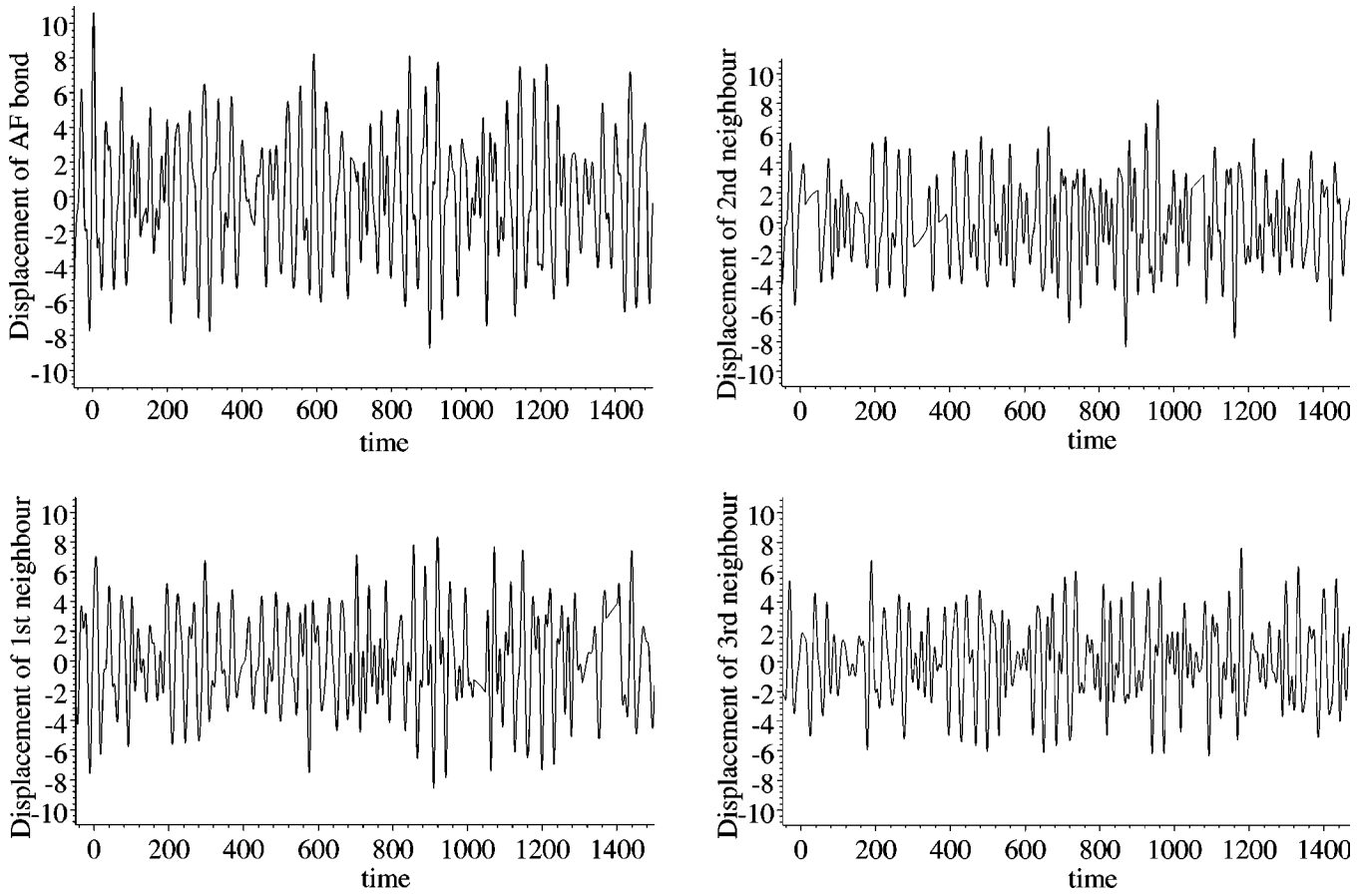

FIG. 6. Plots of the base opening distance against time for the defective base pair (top left), the nearest neighbor ($n=1$, bottom left), the second neighbor ($n=2$, top right), and the third neighbor (top right), showing the decrease in amplitude as one moves away from the defect. The vertical axis records distance in Å , and the horizontal axis shows time in units of 0.7 ps; the parameters γ,k,ε are as in Fig. 5.

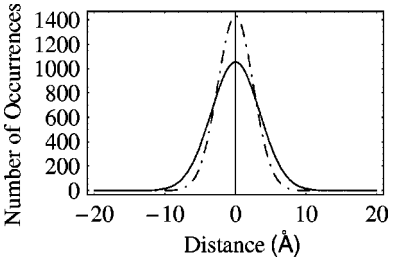

FIG. 7. Distribution of opening distance for the AF base (filled line) and a standard base (dashed line), measured over 10 ns.

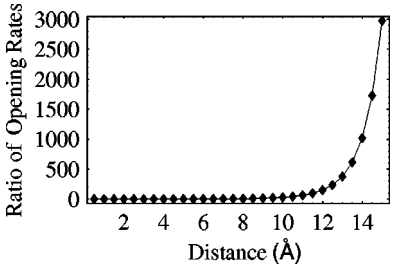

FIG. 8. Ratio of breathing rates for the standard base relative to an AF base. The AF base is expected to breathe around 1000 times faster than the conventional base.

IV. ANALYSIS OF RESULTS

In order to test the model presented here, we have investigated its ability to (a) reproduce the frequency of breathing of an AF base pair as calculated from the atomistic simulations, (b) reproduce the relative frequencies of breathing of a normal AT base pair compared to an AF pair, and (c) reproduce the free energy of breathing for the AF pair, as previously estimated from the atomistic simulations. For this purpose, the model was used to create a 10-ns dynamics simulation, during which the displacement from equilibrium of the variables $y_n(t)$ was sampled every picosecond. Figure 6 illustrates the first ns of the simulation for base pairs $n=0,1,2,3$. The resulting distribution of opening distances is shown in Fig. 7, which contains the results for both the central AF base pair and one of the (standard) base pairs far away from the defect.

In the atomistic simulation [1], breathing was defined in terms of distances between hydrogen-bonded atoms in the AF pair exceeding a certain critical value. With that definition, a breather was observed four times in 10 ns. From the histogram, we have to define a breathing event in this model as one that causes the distance between the bases to exceed 11.5 Å . We now consider whether such a definition of ''breathing'' is consistent with the predicted ratio of breathing frequencies for normal and AF pairs, and for the calculated free energy of breathing of an AF pair.

If we take 11.5 Å to define a breathing event, then in fact within the time scale of our simulation the normal base pairs did not breathe at all. However, it is possible to infer from the fitted Gaussian distribution that such an event should occur once every 250 ns, i.e., a rate of 4 per μs. Data from NMR on imino-proton exchange and chemical shift measurements indicate that for conventional DNA sequences, breathing is an event that takes place on the microsecond time scale, depending on experimental conditions [13]. More precisely, taking 11.5 Å to define base pair breathing, the ratio of opening frequencies for AF and normal base pairs is predicted to be about 100. This may be considered to be at the lower limit of the expected ratio. We have therefore calculated what the opening distance would have to be defined as in order to produce a breathing ratio AF:normal of 1000. This can be calculated from the histograms, and Fig. 8 shows how the ratio varies with the defining opening distance. We see that for a ratio of 1000, we require breathing to be defined as a displacement from equilibrium of 14 Å .

From the atomic simulation, the free energy of opening of the AF base pair was calculated, from the log of the ratio of the time spent open to the time spent not open, as about 1 Kcal/mol. The same procedure may be used here. The calculated free energy of breathing, for AF and normal base pairs, is plotted as a function of the defining opening distance in Fig. 9. It can be seen that if we define opening as displacements above 11.5 Å , the free energy of breathing for the AF pair is defined as 4.2 Kcal/mol. If, on the other hand, we aim to reproduce the previous estimate of the free energy of breathing, we must define breathing as opening above 4.8 Å. The above results show a remarkable agreement given that only three parameters have been fitted to external data $(k/m,\gamma/m,\varepsilon)$, as described in Sec. II A. Due to the rarity of the breathing event, which is extremely infrequent even in the case of the AF base pair, the results described in this section are very sensitive to the shape of the tail of the distributions displayed in Fig. 7. Here, we have fitted the data to the obvious first choice distribution, namely a Gaussian. However, another distribution may give more accurate results.

V. CONCLUSIONS

We have shown that the presence of an inhomogeneity or defect in the interchain bonds is sufficient to cause a breatherlike mode to exist. Our model has no nonlinearity, which is usually necessary for breathers to exist. Being a linear model, a normal mode decomposition is possible, and it is possible to explicitly find the full set of normal modes using asymptotic analysis, which relies on the interchain bonding being much weaker than the along-chain interactions. This result is interesting in its own right, since localized modes are usually attributed to nonlinearities, and we have shown that they can actually be accounted for more simply by inhomogeneities in the lattice. This does not discount breather

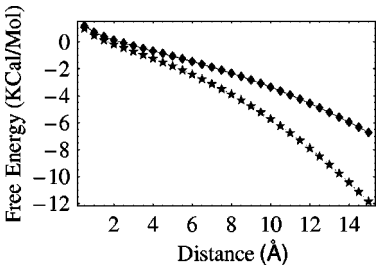

FIG. 9. Free energy associated with an opening event for the AF base (diamonds) and a conventional base (stars).

modes, since traveling localized modes cannot be explained by our analysis.

The model has then been analyzed using the ideas of statistical mechanics to investigate whether the breathing motion is observable when all modes are excited. At leading order, all base pairs have the same mean-square displacement from their equilibrium position, but the first correction term to this result shows a clear peak in displacement at the site of the defect. Figure 6 shows the interpair distance plotted against time for the defective base pair and for the three base pairs closest to it on one side. These show an increased deviation for the defective pair and its nearest neighbor, but little effect on base pairs further away.

Here we have implemented periodic boundary conditions, but other boundary conditions could be used, and we believe that similar results would be seen. Boundary conditions of some description need to be applied so that the correct number of normal modes are found and that the large-N limit is taken correctly, that is, without the introduction of extra, spurious modes that might invalidate subsequent analysis. A hint of this can be seen in the slight difference between the breatherlike mode in the finite system and the infinite system.

In Sec. III only the variables describing the spacing between the chains supported a breather-type solution; the average displacement of the double chain showed no localized oscillation. Experiments in which breathers were observed in only one strand and not the other cannot be explained by the analysis described here; for that a more general model is required, which we hope to present later.

The comparison with results of molecular simulations detailed in Sec. IV shows that defining a displacement of more than 11 Å to be a breathing event gives a reasonable time scale for the occurrences, both for the defective AF base pair and for a standard base pair. This also gives free energy of breathing, which is of the correct order of magnitude. For such a simple, generic model, in which only three parameters are fitted to known data, these results show a remarkably good accuracy.

ACKNOWLEDGMENT

We wish to thank Dr. Keith Benedict for helpful comments on the manuscript.

APPENDIX: CALCULATIONS

1. Calculation of $j_n^{(r)}$

We rewrite I_n using trigonometric identities as

$$I_n=\int_0^1 \frac{\sin^2(n\mu\pi)\ \ d\mu}{q+\sin^2(\frac{1}{2}\mu\pi)}=\int_0^1 \frac{[1-\cos(2n\pi\mu)]\,d\mu}{2q+1-\cos(\pi\mu)} \tag{A1}$$

and use $1+2q=\sqrt{1+\tilde{q}^2}$. Now define

$$j_n^{(r)}=\int_0^1 \frac{\cos(n\pi d\mu)}{\sqrt{1+\tilde{q}^2}-\cos(\pi\mu)},$$

$$j_n^{(i)}=\int_0^1 \frac{\sin(n\pi d\mu)}{\sqrt{1+\tilde{q}^2}-\cos(\pi\mu)} \tag{A2}$$

so that $I_n=j_0^{(r)}-j_{2n}^{(r)}$ and

$$j_n=j_n^{(r)}+ij_n^{(i)}=\int_0^1 \frac{e^{n\pi i\mu}}{\sqrt{1+\tilde{q}^2}-\cos(\pi\mu)}. \tag{A3}$$

We then define $J(\delta)=\Sigma_{n=0}^{\infty}\delta^n j_n$ and swap the order of the summation and the integration. Performing the summation inside the integral yields

$$J(\delta)=\int_0^1 \frac{1-\delta\cos(\pi\mu)+i\,\delta\sin(\pi\mu)\ \ d\mu}{(\sqrt{1+\tilde{q}^2}-\cos\pi\mu)(1+\delta^2-2\,\delta\cos\pi\mu)}, \tag{A4}$$

from which it can be shown that

$$J^{(r)}(\delta)=\mathrm{Re}\{J(\delta)\}=\frac{1-\delta\tilde{q}-\delta\sqrt{1+\tilde{q}^2}}{\tilde{q}(1+\delta^2-2\,\delta\sqrt{1+\tilde{q}^2})}. \tag{A5}$$

Since $J^{(r)}(\delta)=\Sigma_{n=0}^{\infty}\delta^n j_n^{(r)}$, we can find the $j_n^{(r)}$ by equating the coefficients of powers of δ in

$$\tilde{q}(1+\delta^2-2\,\delta\sqrt{1+\tilde{q}^2})J^{(r)}(\delta)=1-\delta\tilde{q}-\delta\sqrt{1+\tilde{q}^2}. \tag{A6}$$

This yields the recurrence relation

$$0=j_n^{(r)}+j_{n-2}^{(r)}-2j_{n-1}^{(r)}\sqrt{1+\tilde{q}^2}, \tag{A7}$$

which has the general solution

$$j_n^{(r)}=C_1(\sqrt{1+\tilde{q}^2}+\tilde{q})^n+C_2(\sqrt{1+\tilde{q}^2}-\tilde{q})^n. \tag{A8}$$

The constants C_1 and C_2 are determined by equating the $O(\delta^0)$ and $O(\delta^1)$ terms, which are $\tilde{q}j_0^{(r)}=1$ and $\tilde{q}(j_1^{(r)}-2j_0^{(r)}\sqrt{1+\tilde{q}^2})=-\tilde{q}-\sqrt{1+\tilde{q}^2}$. These equations imply $C_1=0$ and $C_2=1/\tilde{q}$. This is consistent with $j_n^{(r)}\to 0$ as $n\to\infty$, as we expect in this limit, due to the rapid oscillations in the integrand of $j_n^{(r)}$ for large n. Thus

$$j_n^{(r)}=\frac{(\sqrt{1+\tilde{q}^2}-\tilde{q})^n}{\tilde{q}}\sim\frac{e^{-n\tilde{q}}}{\tilde{q}}. \tag{A9}$$

Rewriting in terms of our original variables, we find $\tilde{q}^2=4q(1+q)$ and hence

$$I_n=\frac{1}{2\sqrt{q(1+q)}}\{1-[1+2q-2\sqrt{q(1+q)}]^{2n}\}$$

$$\sim\frac{1}{2\sqrt{q}}[1-\exp(-4n\sqrt{q})], \tag{A10}$$

since we are interested in the limit $n\to\infty$ with $q\to 0$.

2. Calculation of $\eta_n^{(i)}$

Here we tackle a similar integral using similar methods. We wish to calculate

$$\eta_n^{(i)}:=\int_0^1\frac{\sin(n\mu\pi)\ \ d\mu}{\sin(\mu\pi)[\,2q+1-\cos(\mu\pi)\,]}, \tag{A11}$$

thus we also define

$$\eta_n^{(r)}:=\int_0^1\frac{\cos(n\mu\pi)\ \ d\mu}{\sin(\mu\pi)[\,2q+1-\cos(\mu\pi)\,]},\quad \eta_n:=\eta_n^{(r)}+i\,\eta_n^{(i)}=\int_0^1\frac{e^{in\mu\pi}\ \ d\mu}{\sin(n\mu\pi)[\,2q+1-\cos(\mu\pi)\,]}, \tag{A12}$$

and then $H(\delta):=\Sigma_{n=0}^{\infty}\delta^n\eta_n$. Reversing the order of the integration and the summation, we find

$$H(\delta)=\int_0^1\frac{d\mu}{\sin(\mu\pi)[\,2q+1-\cos(\mu\pi)\,][\,1-\delta\cos(\mu\pi)-i\delta\sin(\mu\pi)\,]}. \tag{A13}$$

To calculate $\eta_n^{(i)}$, it is sufficient for us to find $H^{(i)}(\delta)=\mathrm{Im}\{H(\delta)\}$, which can be found using MAPLE,

$$H^{(i)}(\delta)=\frac{\delta(1-4\delta\sqrt{q(1+q)}-\delta^2)}{2\sqrt{q(1+q)}(1-\delta^2)[1-2(1+2q)\delta+\delta^2]}. \tag{A14}$$

Recalling $H^{(i)}=\Sigma_{n=0}^{\infty}\delta^n\eta_n^{(i)}$, we find the $\eta_n^{(i)}$'s by equating coefficients of powers of δ in

$$2\sqrt{q(1+q)}(1-\delta^2)[1-2(1+2q)\delta+\delta^2]\sum_{n=0}^{\infty}\delta^n\eta_n^{(i)}$$

$$=\delta(1-4\delta\sqrt{q(1+q)}-\delta^2). \tag{A15}$$

For δ^n with $n\geqslant 4$, we find the recurrence relation

$$\eta_{n+4}^{(i)}-2(1+2q)\,\eta_{n+3}^{(i)}+2(1+2q)\,\eta_{n+1}^{(i)}-\eta_n^{(i)}=0, \tag{A16}$$

which has the solution

$$\eta_n^{(i)}=C_0+C_1(-1)^n+C_2\lambda^n+C_3\lambda^{-n}, \tag{A17}$$

with $\lambda=1+2q+2\sqrt{q(1+q)}$. Powers $n=0,1,2,3$ yield

$$\eta_0^{(i)}=0,\quad \eta_1^{(i)}=\frac{1}{2\sqrt{q(1+q)}}, \tag{A18}$$

$$\eta_2^{(i)}=\frac{1+2q-2\sqrt{q(1+q)}}{\sqrt{q(1+q)}},$$

$$\eta_3^{(i)}=\frac{4(1+2q)^2-8(1+2q)\sqrt{q(1+q)}-1}{2\sqrt{q(1+q)}}.$$

Applying these to the solution (A17) yields the constants C_k $(k=0,1,2,3)$ and hence the final solution

$$\eta_n^{(i)}=\frac{\lambda}{(\lambda^2-1)^2}[(\lambda+1)^2-(\lambda-1)^2(-1)^n-4\lambda^{1-n}]. \tag{A19}$$

This simplifies for the case of even n, which we require, and simplifies further in the limit of small q,

$$\eta_{2n}^{(i)}=\frac{4\lambda^2(1-\lambda^{-2n})}{(\lambda^2-1)^2}=\frac{1-[1+2q-2\sqrt{q(1+q)}]^{2n}}{4q(1+q)}$$

$$\sim\frac{1-e^{-4n\sqrt{q}}}{4q}. \tag{A20}$$

[1] E. Cubero, E.C. Sherer, F.J. Luque, M. Orozco, and C.A. Laughton, J. Am. Chem. Soc. **121**, 8653 (1999).
[2] L.V. Yakushevich, Physica D **79**, 77 (1994).
[3] T. Lipniacki, Phys. Rev. E **60**, 7253 (1999).
[4] J.A.D. Wattis, J. Phys. A **31**, 3301 (1998).
[5] S. Cocco and R. Monasson, J. Chem. Phys. **112**, 10 017 (2000).
[6] M. Peyrard and A.R. Bishop, Phys. Rev. Lett. **62**, 2755 (1989).
[7] M. Peyrard and A.R. Bishop, Lecture Notes in Physics, Vol. 353, edited by M. Barthes and J. Leon (Springer, Berlin, 1990), pp. 29–41.
[8] M. Salerno, Phys. Rev. A **44**, 5292 (1991).
[9] J.J.L. Ting, Int. J. Mod. Phys. A **7**, 1125 (1998).
[10] W.K. Olson and V.B. Zhurkin, Curr. Opin. Struct. Biol. **10**, 286 (2000).
[11] W.K. Olsen, N.L. Marky, R.L. Jernigan, and V.B. Zhurkin, J. Mol. Biol. **232**, 530 (1993).
[12] Y.Z. Chen, V. Mohan, and R.H. Griffey, Phys. Rev. E **61**, 5640 (2000).
[13] L.V. Yakushevich, *Nonlinear Physics of DNA* (Wiley, Chichester, 1998).