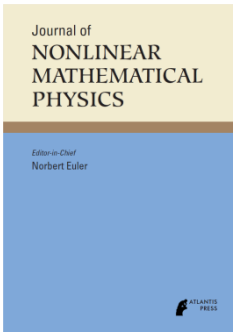


ATLANTIS PRESS

Journal of Nonlinear Mathematical Physics

ISSN (Online): 1776-0852 **ISSN (Print): 1402-9251**
Journal Home Page: https://www.atlantis-press.com/journals/jnmp

Soliton Propagation in Homogeneous and Inhomogeneous Models for DNA Torsion Dynamics

Mariano Cadoni, Roberto de Leo, Sergio Demelio

To cite this article: Mariano Cadoni, Roberto de Leo, Sergio Demelio (2011) Soliton Propagation in Homogeneous and Inhomogeneous Models for DNA Torsion Dynamics, Journal of Nonlinear Mathematical Physics 18:Supplement 2, 287–319, DOI: https://doi.org/10.1142/S1402925111001544

To link to this article: https://doi.org/10.1142/S1402925111001544

Published online: 04 January 2021

ARTICLE

World Scientific
www.worldscientific.com

ATLANTIS PRESS

Journal of Nonlinear Mathematical Physics, Vol. 18, Suppl. 2 (2011) 287–319

DOI: 10.1142/S1402925111001544

SOLITON PROPAGATION IN HOMOGENEOUS AND INHOMOGENEOUS MODELS FOR DNA TORSION DYNAMICS

MARIANO CADONI*, ROBERTO DE LEO† and SERGIO DEMELIO‡

Dipartimento di Fisica, Università di Cagliari and INFN
Sezione di Cagliari, Cittadella Universitaria
09042 Monserrato (Italy)
**mariano.cadoni@ca.infn.it*
†*roberto.deleo@ca.infn.it*
‡*sergio.demelio@ca.infn.it*

Received 20 September 2010
Revised 7 February 2011
Accepted 11 February 2011

The existence of solitonic excitations is a generic feature of a broad class of homogeneous models for nonlinear DNA internal torsional dynamics, but many properties of solitonic propagation depend on the actual model one is considering. In this paper we perform a detailed and comparative numerical investigation of the profiles and time evolution of solitons for two different models, the Yakushevich one and the more recent "composite" model of [1], and for two different choices of the potential describing the pairing interaction between bases (harmonic and Morse potential). We consider not only homogeneous DNA chains but also inhomogeneous ones (with sequence of bases corresponding to a real organism, the Human Adenovirus 2). We show that twist solitons can propagate in inhomogeneous chains over biologically significant distances. It is also shown that stable soliton propagation is possible for inhomogeneous chains when dissipation and an external force are present. On a more general level, our results indicate that solitonic propagation can take place in highly inhomogeneous nonlinear media.

Keywords: Nonlinear mechanical models of DNA; soliton propagation in DNA chains.

1. Introduction

The idea that solitons could play a functional role in DNA dynamics, in particular in transcription and denaturation, goes back to early works on solitons in biological systems [2, 3] and is attractive in many ways. Thus the study of nonlinear DNA dynamics by means of mesoscopic models [4–8] did to a large extent focus on breathers, domain walls and soliton solutions.

On the other hand, this conjecture is subject to several solid objections both from the biological point of view and from the point of view of nonlinear dynamics. In fact, the mesoscopic models usually considered are: (i) homogeneous, hence disregarding the differences between the bases which embody the genetical information; (ii) Hamiltonian, hence conservative and disregard dissipation due to interactions with the DNA environment and

external forcing; (iii) considering DNA *per se*, hence disregarding not only interaction with its environment, as already mentioned above, but also interaction with other molecules playing a substantial role in biological process, e.g. RNA Polymerase (RNAP) in transcription (iiii) a rough simplified model of real DNA because they use a bunch of degrees of freedom (usually two or four) to describe each base pair, which contains about 200 degrees of freedom (DOF). It is generally believed that renouncing to any of these features would produce a model in which nonlinear excitations of solitonic nature could not propagate over biologically significant distances, in particular for models of DNA internal torsion dynamics such as the Yakushevich model and related ones [4].[a]

On the other hand, perturbation analysis of a recently introduced "composite" model [1, 11–13] suggests that the latter could be able to support solitons with a significantly long life also for inhomogeneous chains, thus providing a way to remove limitation (i) above. The aim of this paper is to study this question beyond the perturbation approach — which is of course possible only numerically. We will find that indeed the inhomogeneous version of such a "composite" model supports solitons able to travel over a biologically significant distance. We will, for the sake of completeness, study both "simple" and "composite" versions of the model (the "simple" version being just the classical Yakushevich model), and different choices for the pairing potential (the harmonic potential first considered by Yakushevich and a Morse potential). We anticipate that — not surprisingly — the Morse potential yields physically better results. Throughout the paper we refer to these models with Y for the Yakushevich model, CY for the composite Yakushevich model and with YM and CYM for the respective models with the harmonic potential replaced by a Morse one.

We will also look at limitation (ii), i.e. consider a version of the (homogeneous and inhomogeneous) models in which one introduces — in the simplest manner — dissipative effects originating in interaction with the fluid environment, and external forcing. Here again we draw inspiration on work by Yakushevich [14], we find that stable soliton propagation in the inhomogeneous chains is possible even in presence of dissipative effects if an external force is also present. Again we get better results by using the Morse pairing potential.

In this paper we will not address the problems related with the limitation (iiii) — we will still work in the framework of a mesoscopic model which uses four DOF to model each base pair. The existence of solitons in a model taking fully into account the 200 DOF of a single base pair remains therefore an open question, which has to be approached with single molecule computational methods.

When considering inhomogeneous models, we will routinely use the base sequence corresponding to a real organism, i.e. the Human Adenovirus 2 (HA2); we will often refer to such a model as "a real DNA model" for short.

It is natural to ask about correspondence of the numerical experiments to be discussed below with real experiments, in particular — as our model investigates the behavior of a single DNA molecule — *single molecule experiments* (SME) [15, 16]. In recent years SME

[a]Breather solutions are more robust, and they have been widely studied in inhomogeneous chains, e.g. via the Peyrard–Bishop–Dauxois model and related ones [5, 9]. Also, it is known that this class of models — which deals with internal stretch rather than torsion dynamics — can support travelling solitons even in the presence of inhomogeneities [10].

have been quite successful in several aspects, and in particular in analyzing DNA elastic properties (also involving global DNA torsion dynamics) and structural transitions [17–30]; SME can also be performed for DNA unzipping, and actually the base sequence can be reconstructed from such measurements [31]; similarly, base stacking forces can be directly measured in SME [32]. SME testing the Peyrard–Bishop–Dauxois model [9] have been recently proposed [33].

As mentioned above, the DNA torsion analyzed by SME is that of the two strands as a whole, i.e. *global* DNA torsion, while in the present paper we consider *internal* DNA torsion, i.e. torsional movements at the level of single nucleotides or even their components. At present, SME are not able to resolve internal DNA dynamics in such detail.

In the context of DNA transcription, a process for which it is conjectured that torsional solitons could accompany the RNAP motion, it is possible to track the movement of RNAP and the global DNA rotation [34]; one observes a behavior compatible with the hypothesis that RNAP moves at nearly constant speed between stops[b] (see also [35–39]). We are not aware of any experiment directly investigating the presence of solitons in DNA with no RNAP attached.[c]

On a more general level our paper can be also considered as a purely theoretical nonlinear science study of soliton propagation in homogeneous and inhomogeneous molecular chains described by models inspired by DNA torsional dynamics. From this general point of view, the main outcome of our investigation is that solitonic propagation can take place also in inhomogeneous molecular chains — and more generally in highly inhomogeneous nonlinear media.

The structure of the paper is as follows. In Sec. 2, we present the inhomogeneous, discrete, version of the CY model for DNA torsional dynamics, and derive the Y model as limiting case of the model. In Sec. 3, we briefly discuss the physical parameters entering in the model.[d] In Sec. 4, we illustrate the algorithms we use for the evaluation of the initial profiles and time evolution of the solitons. In Sec. 5, we present the results for the time evolution of solitons in a homogeneous DNA in the case of the Y, C, YM and CYM models. In Sec. 6, we present our main results, i.e. those regarding the time evolution of solitons in inhomogeneous chains. Here we consider both a real DNA sequence (the Human Adenovirus 2) and a purely random DNA sequence. Again, we investigate the cases of the four models Y, C, YM and CYM. The discussion up to this point considers DNA as an isolated, conservative system. In Sec. 7, we take into account the effect of dissipation and of external forces, thus inserting the interaction of DNA with its environment in our model. Our approach follows the recent work by Yakushevich [14] (see [41] for a different approach), and fully confirms her result concerning the homogeneous Y model; here we also study the inhomogeneous Y model, as well as (homogeneous and inhomogeneous versions of) the YM model, deferring consideration of the CY and CYM models with dissipation and external forces to later work. The final Sec. 9 summarizes our discussion and results.

[b]This is necessary if RNAP motion is related to a soliton-like excitation traveling along the DNA chain.
[c]This appears to be still beyond our technical possibilities. It was proposed by Yakushevich to use neutron scattering to this aim, but the experiment was never realized.
[d]All the estimates of the geometrical and dynamical parameters included in this paper are an improvement and/or an update of the corresponding ones published in [1] and [40].

2. The Inhomogeneous Version of the Composite Yakushevich Model for DNA Torsional Dynamics

Following the characterization of relevant internal DOF of the DNA chain, two classes of mesoscopic model, aiming to describe the DNA internal dynamics, have been extensively studied in the Nonlinear Physics literature: the "radial" model by Peyrard and Bishop [42] and its extensions [5, 9, 43–46]; and the "torsional" model by Yakushevich [4, 47] and its extensions [1, 48–52] (see [41, 53–55] for more recent developments). In this paper we will focus on torsional models for DNA dynamics.

The simplest mesoscopic model describing the dynamics of DNA rotational modes, was proposed by Yakushevich [4, 47] (see [4] for previous similar models) and considers a single torsional DOF per nucleotide; this is a homogeneous model and supports sine-Gordon (SG) solitons. Some time ago a CY model generalizing the Y model was proposed by some of us [1] (see also [11, 13, 40]). This considers that internal torsional motions can take place both as a rotation of the nitrogen base with respect to the sugar ring and as a rotation of the sugar-phosphate group [56, 57]. Thus the "composite" torsional model describes the state of each nucleotide by *two* independent angular DOF (one related to the sugar-phosphate group and one to the nitrogen base). Modeling the DNA as a double chain of double rotators gives room to a somewhat more detailed description of both the DNA geometry and its internal interactions, and turned out to solve some problem of the Yakushevich model [1].

The CY model for DNA is a double chain of some large but finite number N of coupled double pendula. It is a natural generalization of the well-known model by Yakushevich where, at every node of each chain, the whole group base-sugar-phosphate is represented by a single disc centered at the chain's backbone axis [58].

In this model we split the group in two components represented in concrete by two discs: one again centered about the backbone axis and representing the sugar-phosphate group and the other, representing the base, which can rotate about a fixed point on the sugar as shown in Fig. 1. Note that the genetic information encoded in DNA's molecule is entirely contained in the sequence of bases while the sugar-phosphate backbone is homogeneous, so the increase we introduce in the geometric detail of the model is rather naturally suggested by the general principle to treat separately the homogeneous and inhomogeneous components of the chain.

In Fig. 1, we show in detail the structure of the model (left) and the geometry of a single node (right). The coordinates of points A, B, C and D, with respect to a coordinate system centered in DNA's symmetry axis and with the x axis directed along the centers of the sugars, are given by the following:

$$
\begin{aligned}
A_{n,i} &= ((-1)^i h_{n,i}, 0, n\delta),\\
B_{n,i} &= A_{n,i} + R((-1)^{i+1}\cos\theta_{n,i}, (-1)^{i+1}\sin\theta_{n,i}, 0),\\
C_{n,i} &= B_{n,i} + (d_{bs} + r_{n,i})(\cos(\theta_{n,i} + \varphi_{n,i}), \sin(\theta_{n,i} + \varphi_{n,i}), 0),\\
D_{n,i} &= C_{n,i} + r_{n,i}(\cos(\theta_{n,i} + \varphi_{n,i}), \sin(\theta_{n,i} + \varphi_{n,i}), 0),
\end{aligned}
$$

where $h_{n,i} = R + d_{bs} + 2r_{n,i} + d_{eq}/2$, $n = 1, \ldots, N$, $i = 1, 2$, d_{eq} is the equilibrium distance between the bases, d_{bs} is the distance of the bases from the sugar and δ is the distance between sites along the chain.

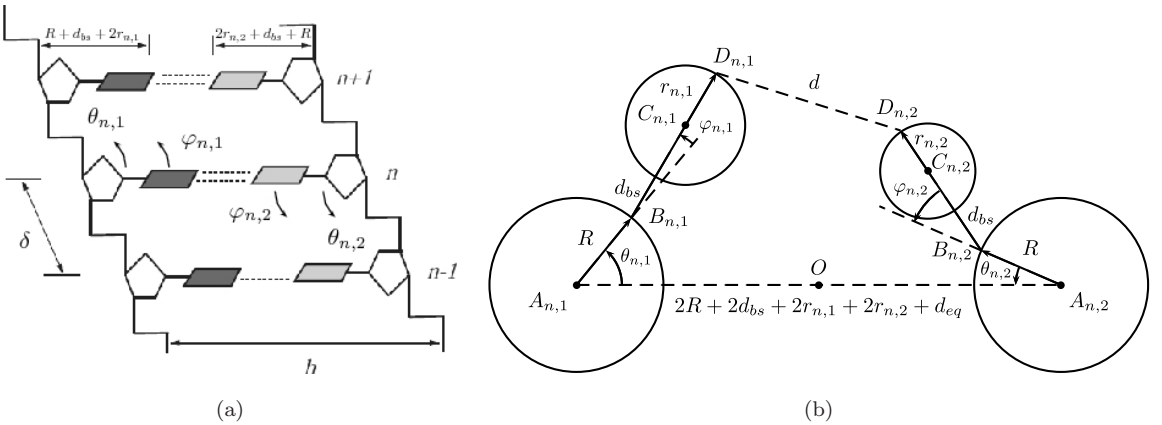


Fig. 1. (a) (adapted from [58]) A fragment of the nonhomogeneous CY model of the DNA's double chain: at every node of the chain there are four degrees of freedom, two for each sugar-phosphate group (the $\{\theta_{n,i}\}$) and two for each base (the $\{\varphi_{n,i}\}$). (b) Detail of a chain node. Every base $\{n, i\}$ is allowed to rotate about the atom of the corresponding sugar, represented by the point $B_{n,i}$, only by an angle between $[-\phi_0, \phi_0]$, with $\phi_0 < \pi$, because of the physical constraint represented by the sugar pentagon. This steric constraint is implemented in the model dynamically through an effective potential. See Sec. 3 for an evaluation of all geometrical constants that appear in the pictures above.

From the mechanical point of view our DNA model is a conservative system with a $4N$-dimensional torus with coordinates $(\theta_{n,i}, \varphi_{n,i})$, $n = 1, \ldots, N$, $i = 1, 2$. The bases though cannot make a complete rotation about the sugar-phosphate group because they would collapse on it (see Fig. 1) and electrostatic forces would not allow this. We model this fact through an effective confining potential V_c whose energy wall is high enough to restrict the range of the bases' angles $\varphi_{n,i}$ to some segment $I = [-\phi_0, \phi_0]$, $\phi_0 < \pi$; for this reason we call the sugar-phosphate angles *topological* and the bases angles *nontopological*.

The kinetic and potential energies of the system are given by the following quantities:

(1) The kinetic energy T_t of the sugar-phosphate group:

$$T_t = \sum_{n=1}^{N} \sum_{i=1}^{2} \frac{1}{2} m_s \dot{B}_{n,i}^2 = \sum_{n=1}^{N} \sum_{i=1}^{2} \frac{1}{2} I_t \dot{\theta}_{n,i}^2, \tag{2.1}$$

where m_s is the mass of the sugar ring and $I_t = m_s R^2$.

(2) The kinetic energy T_s of the bases:

$$T_s = \sum_{n=1}^{N} \sum_{i=1}^{2} \frac{1}{2} m_{b_{n,i}} \dot{C}_{n,i}^2 = \sum_{n=1}^{N} \sum_{i=1}^{2} \frac{1}{2} I_{s_{n,i}} [\delta_{n,i}^2 \dot{\varphi}_{n,i}^2 + 2\delta_{n,i}(\delta_{n,i} + \alpha \cos \varphi_{n,i}) \dot{\varphi}_{n,i} \dot{\theta}_{n,i} + (\delta_{n,i}^2 + 2\alpha \delta_{n,i} \cos \varphi_{n,i} + \alpha^2) \dot{\theta}_{n,i}^2], \tag{2.2}$$

where $b_{n,i}$ is the sequence of bases in the chain, $m_{b_{n,i}}$ their masses, $I_{s_{n,i}} = m_{b_{n,i}} r_{n,i}^2$ their inertia momenta with respect to the points $B_{i,n}$, $\delta_{n,i} = (d_{bs} + r_{n,i})/(d_{bs} + \bar{r})$ and $\alpha = R/(d_{bs} + \bar{r})$ ($\bar{r}$ is the mean value of the r_{ni}'s, see below). Notice, however, that in our calculations we evaluate the inertia momenta of the bases directly out of DNA's molecular structure in order to obtain more reliable values for them.

(3) The *torsional* potential energy V_t, modelling the interaction between next-neighbor sugar-phosphate groups on the same chain, i.e. the torsional elasticity of the backbone. This force is the result of complex molecular interactions at the backbone level and it has to be considered an "effective" interaction term. In order to keep the potential expression as simple as possible, we use for it the "physical pendulum" periodic potential:

$$V_t = \sum_{n=1}^{N-1} \sum_{i=1}^{2} K_t[1 - \cos(\Delta\theta_{n,i})], \tag{2.3}$$

where $\Delta\theta_{n,i} = \theta_{n+1,i} - \theta_{n,i}$ and K_t is the torsional coupling constant.

(4) The *stacking* potential models the $\pi - \pi$ bonds between the rings that constitute the bases. This interaction is much better understood than the previous one and in particular it is clear that it only depends on the relative displacement between next-neighbor bases, going rapidly to zero together with the overlapping portion of their surface, e.g. like in a Morse-like potential. For simplicity we use for it a simple harmonic bond on the "xy" distance between the centers of the bases $d^2_{xy}(C_{n+1,i}; C_{n,i})$. We will disregard in this work all inhomogeneities in this potential and we leave to a future work the study of a fully inhomogeneous model. In particular we consider a mean value $\bar{r}$ for the four values $r_{n,i}$ of the radii of the basis and we set (k_s is the stacking coupling constant)

$$V_s = \sum_{n=1}^{N-1} \sum_{i=1}^{2} \frac{1}{2} k_s \, d^2_{xy}(C_{n+1,i}; C_{n,i}) = \sum_{n=1}^{N-1} \sum_{i=1}^{2} \frac{1}{2} K_s \, l^2_{n,i}, \tag{2.4}$$

where $K_s = k_s(d_{bs} + \bar{r})^2$ and $l^2_{n,i} = d^2_{xy}(C_{n+1,i}; C_{n,i})/(d_{bs} + \bar{r})^2$.

(5) The *pairing* potential V_p models the ionic bonds between base-pairs. This is the best understood force among the ones we are considering and, like in the stacking case, it is known to go rapidly to zero a few Angstrom from the equilibrium position. Following [59], we model this behaviour with a Morse-like potential. For the YM and CYM model we have therefore:

$$V_p = \sum_{n=1}^{N} k_{b_n} \kappa[1 - e^{-a(d_{xy}(D_{n,1}; D_{n,2}) - d_0)}]^2 = \sum_{n=1}^{N} k_{b_n} \kappa[1 - e^{-\mu(l_p - l_0)}]^2, \tag{2.5}$$

where κ is the Morse coupling constant, k_{b_n} is the inhomogeneity factor (equal to 1 for $b_n = A, T$ and to 1.5 for $b_n = G, C$) and μ, l_p, l_0, β and $\gamma_{n,i}$ are the dimensionless quantities corresponding respectively to a, $d_{xy}(D_{n,1}; D_{n,2})$, d_{eq}, R and $(d_{bs} + 2r_{n,i})$ with respect to the reference length $(d_{bs} + 2\bar{r})$. In order to simplify the numerical analysis we use the so-called contact approximation $l_0 = 0$ (see [1, 40] for a justification of this approximation). Furthermore, in order to compare our results with analytical results and with several other numerical results available in literature (e.g. in [14, 58, 60]) we also consider, in the case of the Y and CY models, the harmonic approximation $V_p = \sum_{n=1}^{N} \frac{1}{2} k_{b_n} K_p (d_{bs} + 2\bar{r})^2 (l_p - l_0)^2$ of the previous Morse potential, where $K_p = 2\kappa\mu^2/(d_{bs} + 2\bar{r})^2$.

(6) The *helicoidal* potential V_h models the forces between nucleotides in solution due to the Bernal-Fowler filaments. We will consider only those being on opposite helices at

half-pitch distance, as they are near enough in three-dimensional space due to the double helical geometry, and only those between the sugar-phosphate groups. As the nucleotide moves, the hydrogen bonds in these filaments and those connecting the filaments to the nucleotides are stretched and thus resist differential motions of the two connected nucleotides. Since the pitch of the helix corresponds to 10 bases in the B-DNA equilibrium configuration we set

$$V_h = \sum_{n=1}^{N-5} \sum_{i=1}^{2} K_h[1 - \cos(\theta_{n+5,i+1} - \theta_{n,i})] \tag{2.6}$$

where the sum $i + 1$ is meant *modulo* 2.

(7) The *confining* potential V_c models an "effective" interaction representing the steric constraint of the sugars, which prevents the bases bound to them from doing a complete rotation about them. In order to keep it as simple as possible we use

$$V_c = \sum_{n=1}^{N} \sum_{i=1}^{2} K_c(\sin \varphi_{n,i})^{2M} \tag{2.7}$$

where M is some large integer and the coupling constant K_c must be taken big enough to prevent the bases from passing through a $\pm\pi/2$ barrier but also not so big to interfere too much with the dynamics when the $\varphi_{n,i}$ are closer to the equilibrium position.

2.1. *The Yakushevich model as a particular case of the composite model*

If we set the nontopological angles $\varphi_{n,i}$ identically to zero, so that the bases rotate rigidly with the sugars, the coordinate space reduces to a $2N$-dimensional torus and the energies, disregarding the helicoidal interaction, reduce to the following three terms:

(1) Kinetic energy

$$T = \sum_{n=1}^{N} \sum_{i=1}^{2} \frac{1}{2}[I_t + (\delta_{n,i} + \alpha)^2 I_{s_{n,i}}]\dot{\theta}_{n,i}^2 = \sum_{n=1}^{N} \sum_{i=1}^{2} \frac{1}{2} I_{n,i}^Y \dot{\theta}_{n,i}^2$$

(2) Torsional energy

$$V_t = \sum_{n=1}^{N-1} \sum_{i=1}^{2} [K_t + (1 + \alpha)^2 (d_{bs} + \bar{r})^2 K_s](1 - \cos \Delta\theta_{n,i}) = \sum_{n=1}^{N-1} \sum_{i=1}^{2} \frac{1}{2} K^Y (1 - \cos \Delta\theta_{n,i})$$

(3) Pairing energy

$$V_p = \sum_{n=1}^{N} k_{b_n} D[1 - e^{-\mu\, l_p(\theta_{n,i},0)}] = \sum_{n=1}^{N} k_{b_n} D[1 - e^{-\mu^Y l_p^Y(\theta_{n,i})}]$$

where $\mu^Y = (1 + \beta)\mu$ and $\rho_{n,i} = (\gamma_{n,i} + \beta)/(1 + \beta)$.

This is the YM model. In the harmonic approximation it reduces to the usual Y model, as given e.g. in [58], with $K^Y = 2\kappa(\mu^Y)^2$; in particular the radii of the discs in the Yakushevich model are given by $R_{n,i}^Y = h_{n,i} - d_{eq}/2$. These three models reduce to their homogeneous

counterparts, discussed e.g. in [1, 40, 58], by replacing $r_{n,i}$ with $\bar{r}$, $I_{s_{n,i}}$ with I_s and both $\gamma_{n,i}$ and $\delta_{n,i}$ with 1.

3. Physical Parameters of the Model

The evaluation of the main geometrical and dynamical physical parameters involved in DNA's dynamics is far from trivial because of the complexity of its molecular structure and of the difficulty in making measurements of the mechanical quantities associated to it. In [40] the values of these parameters for the CY model have been discussed in detail; these are summarized in the Tables 1 (kinetic parameters) and 3 (coupling constants). In particular, in Table 3 we show the physical range for both the coupling constants a, K_t, K_s, K_p, κ, K_h and the corresponding dimensionless quantities $\mu = a(d_{bs} + 2\bar{r})$, $g_t = K_t/E_0$, $g_s = K_s(d_{bs} + \bar{r})^2/2/E_0$, $g_p = K_p/E_0$ (in the harmonic case) or $g_p = D/E_0$ (in the Morse case), $g_h = K_h/E_0$, where the reference energy $E_0 = 223\,\mathrm{kJ/mol}$ has been chosen as the energy of the pairing coupling in the harmonic approximation (so that in the harmonic approximation $g_p = 1$). In Table 2 we show the corresponding values induced by those above in the Y and YM models.

There are two major criteria to select a set of coupling constants within the range of physically possible values. One is the compliance with set of dispersion relations associated to the homogeneous system ([40] for more details); in particular the speed of acoustic

Table 1. Values of masses, momenta of inertia and basic geometrical parameters occurring in our CY and CYM models of DNA.

	A	T	G	C	Mean	Sugar
m	134	125	150	110	130	85
I	3.6×10^3	3.0×10^3	4.4×10^3	2.3×10^3	3.3×10^3	2.9×10^3
$l(2\bar{r})$	3.9	2.9	4.1	2.7	3.4	3.1
d_{bs}	1.0	1.0	1.0	1.0	1.0	—
d_{eq}	3.0	3.0	3.0	3.0	3.0	—

Units of measure are: atomic unit for masses m, $1.67\times10^{-47}\mathrm{Kg}\cdot\mathrm{m}^2$ for momenta of inertia I, Angstrom for l (the longitudinal width of bases and sugar), d_{bs} (the distances sugar-base) and d_{eq} (the distance at the equilibrium for the pairs AT and GC). Note in particular that $\bar{r} = \bar{l}/2$. These values have been extracted from the sample "generic" B-DNA PDB data [61], provided by the Glactone Project [62], and agrees within 5% with the data from [63]. The lengths of bases and sugar were taken by projecting them on the direction passing through the two phosphate atoms of the chain node to best fit the geometry of the model; the (real) measure for the diameters of the bases A, T, G, C, and the sugar and for d_{bs} are respectively (in Å): 4.6, 4.0, 5.7, 4.0, 3.3 and 1.5.

Table 2. Masses, momenta of inertia and radii for the Y and YM models, obtained through a dynamical reduction of the values of Table 1.

	A	T	G	C	Mean
m^Y	219	210	235	195	214.75
I^Y	21×10^3	15.6×10^3	25.7×10^3	12.3×10^3	18.6×10^3
R^Y	8	7	8.2	6.8	7.5

Units are atomic units for masses m^Y, $1.67 \times 10^{-47}\,\mathrm{Kg}\cdot\mathrm{m}^2$ for the inertia momenta I^Y, Angstrom for R^Y.

Table 3. (above) Upper and lower bounds of the coupling constants appearing in the CY and CYM models and (below) upper and lower bounds of the corresponding dimensionless quantities (see [40]), defined as $\mu = a(d_{bs} + 2\bar{r})$, $g_t = K_t/E_0$, $g_s = K_s(d_{bs} + \bar{r})^2/2/E_0$, $g_h = K_h/E_0$, $g_p = K_p/E_0$ in the harmonic case and $g_p = D/E_0$ in the Morse case, where $E_0 = K_p(d_{bs} + 2\bar{r})^2/2 \simeq 223\,\mathrm{kJ/mol}$. Note that in the remaining of the paper we use the same symbol g_p for the pairing coupling constant in the cases of harmonic and Morse versions of it; it will be clear from the context which one we refer to.

	K_t	K_s	K_p	κ	a	K_h
lo bd	130 kJ/mol	—	3.5 N/m	30 meV	2 Å^{-1}	$K_t/100$
up bd	720 kJ/mol	16.6 N/m	30 N/m	50 meV	4 Å^{-1}	$K_t/25$
	g_t	g_s	g_p $(harm)$	g_p $(Morse)$	μ	g_h
lo bd	0.58	—	0.91	0.013	8.8	$g_t/100$
up bd	3.2	1.6	7.8	0.02	17.6	$g_t/25$

phonons in DNA is thought to be between 1.9 km/s and 3.5 km/s and the lowest optical frequency ω_0 to be about $35\,\mathrm{cm}^{-1}$. The other is that if we want soliton propagation to model DNA transcription, the width of kinks should be compatible with the diameter of the DNA transcription bubbles.

As often happens, these two requirements are somehow in contrast: in order to have narrow kinks we must keep low K_s and K_t, but this tends to lower the phonons speed below the range above and *vice versa*. Moreover the diameter of a kink[e] is also affected sensibly by its speed, which instead does not play any role in the phonons quantities. Several choices are of course viable. The set used in [40] is not a good choice because induces very wide kink diameters; in this paper we rather use the set shown in Table 4. Observe that we have set the dimensionless confining coupling constant $g_c = K_c/E_0$ to the value $g_c = 10$.

With this choice the velocities of acoustic phonons are $c_2 = \delta\sqrt{(K_t + 25K_h)/I_t} \simeq 1.3\,\mathrm{km/s}$ and $c_3 = \delta\sqrt{K_s/m_b} \simeq 1.3\,\mathrm{km/s}$ and the optical frequency is $q_4 = \sqrt{2K_p/m_b} \simeq 28\,\mathrm{cm}^{-1}$ (the variables names refer to [40]), which are reasonably close to the the physical range mentioned above. On the other hand kinks diameters have a range of variability from 80 bp (base-pairs) at $v = 0$ to 46 bp when the soliton velocity is close to the sound speed.

4. Numerical Analysis of Solitons Motion

We now will investigate, numerically, soliton motion in the models we are considering. This study will involve two independent tasks: first we will determine the initial profile of the

Table 4. Values of the dimensionless dynamical constants chosen as default for the CYM model, where $i_t = \frac{v_0^2}{\delta^2 E_0} I_t$ and $i_s = \frac{v_0^2}{\delta^2 E_0} I_s$ with $v_0 = 1\,\mathrm{km/s}$. After fixing to zero the bases angles we get the YM model with the following values for the corresponding dimensionless parameters: $m = i_t + (1+\alpha)^2 i_s \simeq 7$, $g = 2g_t + 4(1+\alpha)^2 g_s \simeq 7.5$, $K = K_p = 0.02$, $\mu^Y = \mu(1+\beta) \simeq 11$.

i_t	i_s	g_t	g_s	g_p	μ	g_h	g_c
1.1	1.3	1	0.3	0.02	6.3	0.01	10

[e]Throughout the paper we evaluate the diameter of a kink on a chain as the number of nodes close to the center and for which the topological angle is between $\pi/20$ and $2\pi - \pi/20$. The center is defined as the last node such that the sum of energies of all nodes before it is smaller than half of the whole energy.

soliton, which will be the initial condition of the system; and then study how it evolves in time when travelling along the chain. We have performed a systematic numerical analysis of the dynamics of the kink solutions of lowest topological charges (0, 1), (1, 0) and (1, 1) (see below for their definition) of our system with $N = 2000$ bp and $N = 3000$ bp.

4.1. *Evaluation of initial profiles*

Once the dynamical, geometrical parameters and coupling constants are fixed, the profile of the soliton is determined, as an extremum of the action $I = \int L$, by two data: the boundary conditions and the position of its center.

Since we are interested only in kinks, we apply as boundary conditions $\varphi_{1,i} = \varphi_{N,i} = 0$ for the nontopological angles and $\theta_{1,i} = 0$, $\theta_{N,i} = 2k_i\pi$, for some integers k_i (winding — topological — numbers), for the topological ones. We correspondingly use the notation (k_1, k_2) to indicate topological solitons for the angles θ_i. Note that for our purposes we can consider the solution invariant by translations — at least far enough from the boundaries — in the homogeneous approximation while this is far from true in the case of real DNA sequences.

It is very difficult to find exact solutions of our models. In analogy with the continuous case, where solitons solutions $\phi_v(x,t)$ move with constant speed and therefore satisfy the relation $\partial_t\phi_v = -v\partial_x\phi_v$, we look for solutions $q_n(t) = (\theta_{n,1}(t), \theta_{n,2}(t), \varphi_{n,1}(t), \varphi_{n,2}(t))$ that satisfy its discrete analog

$$\dot{q}_n = -v\Delta_n q/\delta. \tag{4.1}$$

Note that the accuracy of this approximation is better when the diameter of the kink is big with respect to the chain step δ and it becomes worse when v approaches the sound speed.

With this ansatz the kinetic energies become

$$\begin{aligned} T_t &= \sum_{n=1}^{N}\sum_{i=1}^{2} \frac{v^2 I_t}{2\delta^2}[\Delta\theta_{n,i}]^2 \\ T_s &= \sum_{n=1}^{N}\sum_{i=1}^{2} \frac{v^2 I_{s_{n,i}}}{2\delta^2}\{\delta_{n,i}^2[\Delta\varphi_{n,i}]^2 + 2\delta_{n,i}(\delta_{n,i} + \alpha\cos\varphi_{n,i})\Delta\varphi_{n,i}\Delta\theta_{n,i} \\ &\quad + (\delta_{n,i}^2 + 2\delta_{n,i}\alpha\cos\varphi_{n,i} + \alpha^2)[\Delta\theta_{n,i}]^2\} \end{aligned} \tag{4.2}$$

and the Lagrangian $L(\dot{\theta}_{n,i}, \dot{\varphi}_{n,i}, \theta_{n,i}, \varphi_{n,i})$ can be written as a sum $L = \sum_{n=1}^{N}\sum_{i=1}^{2} L_{n,i}$, where $L_{n,i} = L(q_n, q_{n+1})$, and therefore it can be seen as a discrete-time action over the circle (see e.g. [64]). The solution of the discrete-time equations are extrema of L and therefore solutions of $dL = (\partial_{\theta_{n,i}}L, \partial_{\varphi_{n,i}}L) = 0$.

We extremize the discrete action with the method of *conjugate gradients* in the independent implementations of NR [65] and GNU's GSL [66], as it has been done in the homogeneous case in [40]; we do not repeat here the discussion and refer the reader to that paper for more details. The CY model is a natural extension of the Y model which, in turn, is a pair of discrete coupled sine-Gordon chains. This is why we tested successfully our code using a simple discrete sine-Gordon chain.

As expected, no kink solutions exist for soliton speeds above a limiting value c depending on all kinetic parameters and coupling constants. This is a generic feature of relativistic

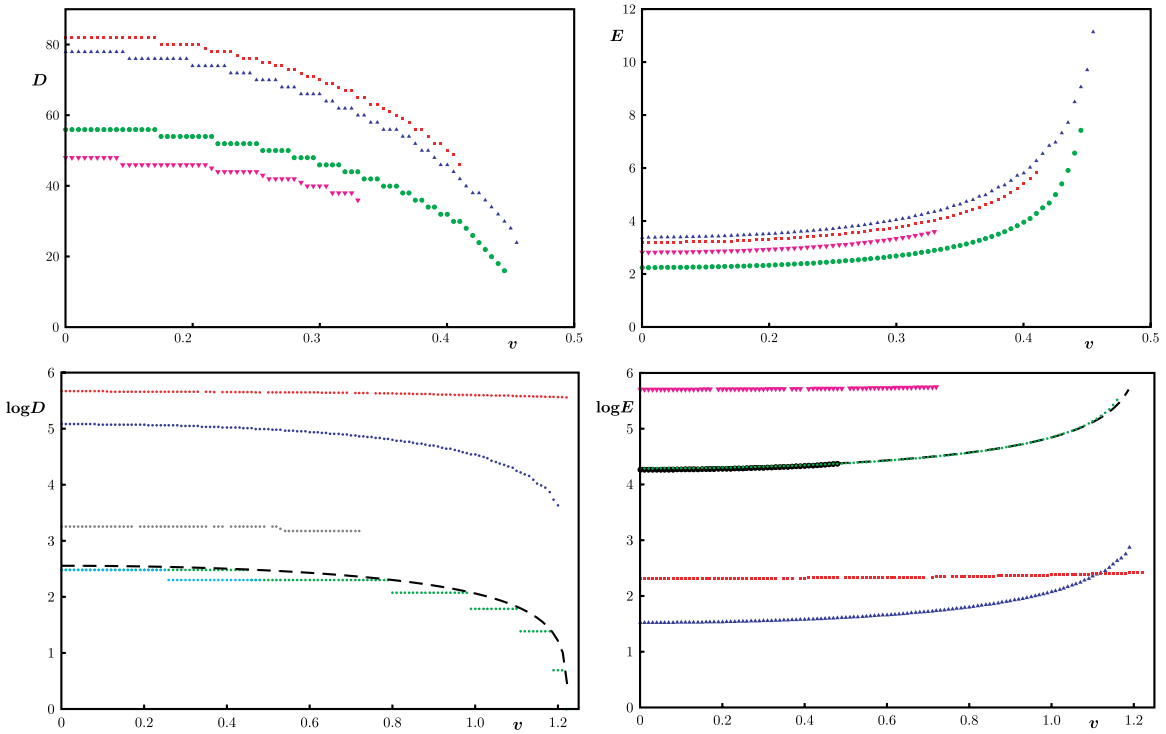


Fig. 2. (above) Plot of kinks' diameters D (left, in bp) and energies E (right) versus their velocity v for the following cases: CYM model, default parameters, topological type $(1,1)$ (squares, red) and $(1,0) \simeq (0,1)$ (∇'s, magenta); YM model, default parameters, topological type $(1,1)$ (Δ's, blue) and $(1,0) \simeq (0,1)$ (circles, green). (below) Logarithmic plot of $(1,1)$ kinks' diameters (left, in bp) and energies (right) versus their velocity parameter v for the following case. We use the default parameters except for the torsion and stacking, which have been increased by a factor 4 to allow the existence of the kinks in the harmonic approximation: SG model (circles, green), Y model (empty circles, grey), CY model (∇'s, magenta), YM model (Δ's, blue) and CYM model (squares, red) — the black dotted line represents the analytical width (left) and energy (right) of the continuous SG kinks. Energies are in units of E_0.

solitons and applies also in the case under consideration. Because of the relativistic behavior we expect the soliton diameters to shrink and their energy to increase when the soliton speed grows. This can be shown to happen for all the models we are considering in this paper. In Fig. 2 (upper row) we show the dependence of energy and diameter of kinks for the lowest topological types $(1,0)$, $(0,1)$ and $(1,1)$ as function of their speed v.[f]

In Fig. 2 (lower row) we show similar results after increasing by a factor 4 the coupling constants of torsion and stacking interactions; after this rescaling, kinks are present in all four models, giving us the opportunity to compare energies and diameters between the models with a Morse potential and those with its harmonic approximation.

4.2. *Time evolution*

The initial data obtained through the previous algorithm represent for us the coordinates $(q_n) = (\theta_{n,1}, \theta_{n,2}, \varphi_{n,1}, \varphi_{n,2})$ at time $t = 0$. Their graph, as function of the discrete variable n,

[f]From now on, by the symbol v we mean the dimensionless ratio between the speed of the soliton and the reference speed $v_0 = 1\,\mathrm{km/s}$. Since there is no ambiguity, we keep calling such v the *velocity* of the soliton.

is the initial profile of a kink with speed v. We recover the initial values of the velocities $\dot{q}_n$ (and therefore of the momenta $p_n = \partial L/\partial\dot{q}_n$) using again (4.1) — we insist that this approximation works well when D/δ is big so we cannot expect a behaviour close to the continuous one for very narrow kinks. Luckily, as we will show below, in our case kinks do not get too narrow.

To study the evolution of the system we were forced to use the Hamiltonian approach rather than the Lagrangian one for the following practical (and unexpected) reason. As shown in Fig. 1, at every node on each chain the geometry of the system is exactly the same of a double pendulum and in particular the coefficients of the kinetic energy are not constant (see Eq. (2.2)). To our knowledge, it seems that there is no numerical algorithm widely available for second order ODEs able to deal with energies whose quadratic form has non-constant coefficients. Hence we were forced to switch to the Hamiltonian formulation, where this complexity is transferred into the Legendre transform between the $\dot{q}_n$ and the p_n and the evolution is represented by a first order ODE.

For the (Hamiltonian) evolution of the system we decided to use two completely different algorithms: the first is the GSL version of the standard Runge–Kutta Prince–Dormand method implemented by M. Galassi *et al.* [66]; the second is a Hamiltonian symplectic integrator implemented and kindly sent to us by its author E. Hairer [67].[g] Several implementations of algorithms based on symplectic integrators are available online, e.g. on the Hairer's website.[h] We made use of them in the Lagrangian (GNI_IRK2, a symplectic and symmetric implicit Runge–Kutta method) and Hamiltonian implementation (GRKAAD). Both codes were provided to us by Hairer in Fortran and translated by us in C/C++.

In principle, the soliton speed v is a free parameter, which can be varied continuously in the range $0 \leq v \leq c$, where c is the limiting velocity. Because we want to investigate the possible role played by solitons in replication process, it will be meaningful to consider in our numerical calculations soliton speeds of the same order of magnitude of the replication process in real DNA, that is $v \sim 1000\,\mathrm{bp/s} \sim 10^{-7}\,\mathrm{m/s}$. However, there is a computational problem that prevent us to perform numerical calculations with soliton speeds of this order of magnitude.

This problem is due to the fact that the natural time scale of our system is of order $10^{-13}\,\mathrm{s}$. This can be easily understood by considering our model as a system of two coupled the Sine–Gordon equations of the form

$$I_t\phi_{tt} - K_t\phi_{nn} + K\sin\phi = 0, \tag{4.3}$$

where I_t are momenta of inertia and K_t, K the torsional and pairing coupling constants and ϕ_{nn} indicates the discretized spatial second derivative. The order of magnitude of the momenta of inertia in play is very small $I_t \simeq 1.3 \cdot 10^{-25}\,\mathrm{kJ\ s^2/mol}$ whereas K_t, K are of

[g]We recall that symplectic integrators are algorithms such that the step function evolution itself is symplectic, i.e. it conserves the symplectic form and therefore the energy; this clearly means that conservation of energy is violated at a much lower rate than nonsymplectic same-order algorithms (see e.g. [67], Chap. 1). Incidentally this fact is very convenient also at the debugging stage, especially when dealing — as we do — with quite complex expressions (the Hamiltonian and its gradient are rather bulky formulas in $4N$ variables), as with such an algorithm even the smallest error in these causes immediate changes in the energy while, when all expressions are correct, energy remains unchanged up to 10^{-8} within (at least) $t < 2000$.

[h]http://www.unige.ch/∼hairer/software.html.

order kJ/mol — see Tables 1–4. If one uses a time unit (TU) of order of magnitude of the second the first term in the right hand side of Eq. (4.3) differs by 25 order of magnitude from the the other two. This scenario is a classical source of errors in numerical analysis, since a huge precision would be needed to evaluate the difference between the two numbers or the multiplication by a factor $\simeq 10^{25}$ would enhance so much the error to make any algorithm useless. On the other hand, using a precision able to deal with this problem would slow down the evolution algorithm so much to make it useless.

The problem is rather solved by defining a a Time Unit $1\,\mathrm{TU} \simeq \delta/v_0 \simeq 3.4 \cdot 10^{-13}\,\mathrm{s}$, so that with this unit all the three terms of Eq. (4.3) are of $\mathcal{O}(1)$. In terms of this TU our reference speed is $v_0 = 1\,\mathrm{km/s} \simeq 1\,\mathrm{bp}/TU$. It follows immediately that our evolution algorithm can be used for soliton speeds of order of magnitude of the limiting speed for our solitons, i.e $v \simeq v_0 = 1\,\mathrm{km/s}$ whereas it becomes useless for soliton speeds of order of the speed of replication process in real DNA, i.e. $v \sim 1000\,\mathrm{bp/s} \sim 10^{-7}\,\mathrm{m/s}$.

For this reason in the next sections of this paper we will mainly consider the time evolution of solitons with velocity v of the order of magnitude of 1 km/s and only in Sec. 8, we will study the evolution of solitons with speed about 5 orders of magnitude less than v_0.

5. Soliton's Motion in Homogeneous DNA Models

DNA is a discrete system and therefore the correct way to describe its dynamics is through the discrete ODE system $d/dt(\partial L/\partial \dot{q}_n) - \partial L/\partial q_n = 0$, rather than its continuous limit for $\delta \to 0$, albeit the latter may be more convenient for analytical discussion. We have thus to discuss solitons in a discrete chain; we will refer to these as *discrete solitons*. In the continuous limit of the homogeneous DNA chain, kinks, if they exist and are stable, propagate at constant speed without loosing energy. We expect this to be not completely true for the discrete homogeneous DNA chain, as invariance under continuous translation is lost. Owing to discreteness of the chain we expect the propagating solitons to loose kinetic energy through phonon emission, although at low rate. Moreover, the phonon emission rate should increase with the soliton speed.

In this section we will investigate in detail the time evolution of discrete solitons for the various DNA models under consideration in the homogeneous approximation, that is all the bases are considered as identical and all the inter-pair hydrogen bonds between bases are taken with the same strength. In the next section we will discuss the general case in which the differences between bases and pairing bonds in a given DNA sequence are taken into account. We will discuss separately the time evolution of solitons under harmonic potential (Subsec. 5.1) and under a Morse potential (Subsec. 5.2), respectively for the Y model (Subsecs. 5.1.1 and 5.2.1) and for the CY model (Subsecs. 5.1.2 and 5.2.2).

5.1. *Time evolution under harmonic pairing potential*

5.1.1. *Yakushevich model*

Motion of kinks in the homogeneous and inhomogeneous Y model has been extensively studied in several papers (see e.g. [58]). Here we study the evolution of Yakushevich kinks as a test for our software and also to produce the profiles with the parameters corresponding — via the dynamical reduction — to those we are using for the CY model; this will allow us to compare the results for the Yakushevich model with those for our generalized model.

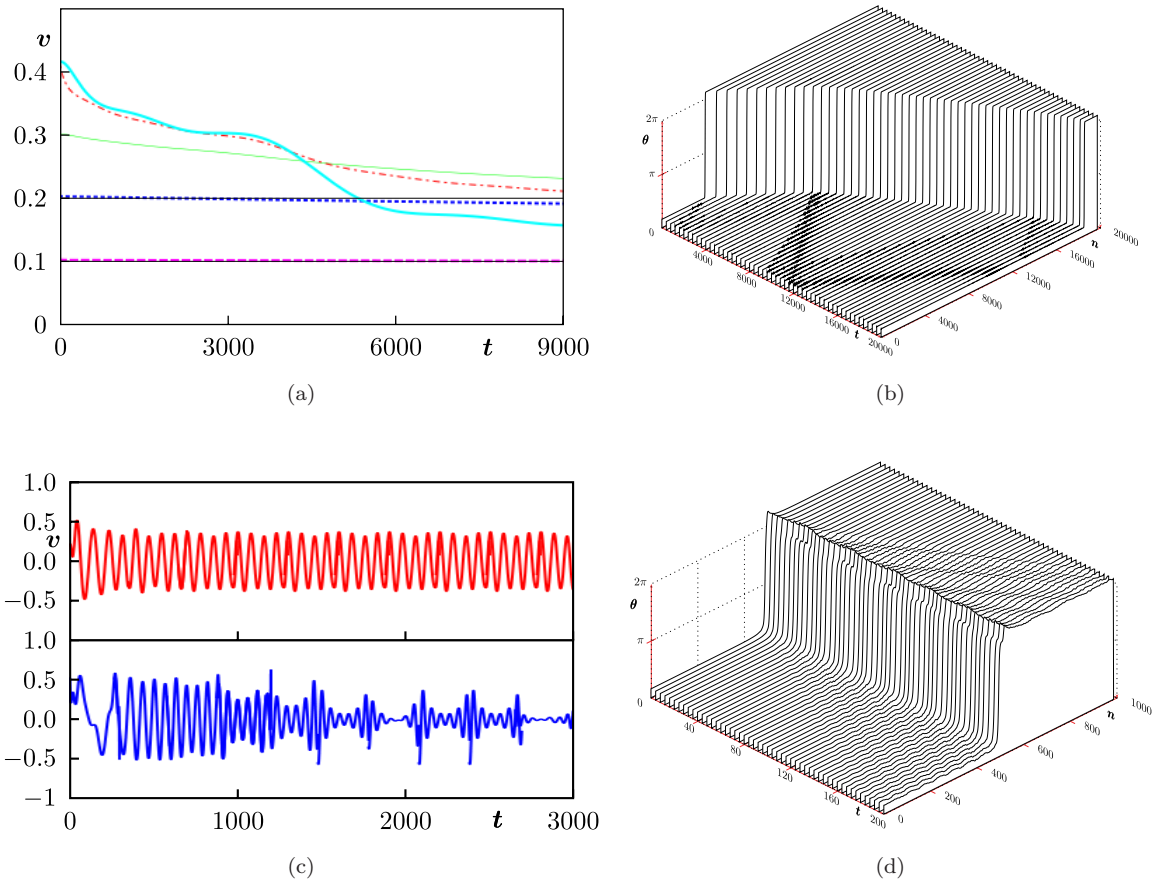


Fig. 3. Time evolution of speeds (in km/s) and profiles of solitons of topological type $(1, 1)$ in the Y model with $m = 7$, $g = 21$, $K = 1$. *Homogeneous chain:* (a) Evolution of the speed of the center of kinks corresponding to the velocity parameters (from bottom to top) $v = 0.1, 0.2, 0.3, 0.4, 0.484 \simeq c$. As expected, when $v \ll c$ the speed of the kink center is close to constant while as v gets close to the limit speed c the kink tends to lose energy and slows down. (b) Evolution of the profile of the kink with $v = 0.484$, showing clearly the deceleration of the kink and the emission of phonons. *HA2 chain:* (c) Evolution of the speed of the center of kinks corresponding to $v = 0.2$ (top) and $v = 0.38 \simeq c$ (bottom). In both cases the kink immediately starts making small oscillations about its initial position. (d) Evolution of the profile of the kink with $v = 0.2$.

In Figs. 3(a) and 3(b), we show the numerical results for the coupling constants $m = 7$, $g = 21$ and $K = 1$, already used in [40] when we discussed the existence of static profiles, and compare them with the corresponding solution in the continuous approximation. It turns out that, when v is not close to the limit speed, the motion is still close to a constant speed one. When v is very close to the limit speed, instead, at the beginning of motion a strong phonon emission takes place, causing a deceleration until the speed reaches some smaller value.

5.1.2. *Composite model*

The time evolution of solitons in the CY model in the harmonic approximation is very simple: solitons do not propagate. In Fig. 5, we show profiles and time evolution of two

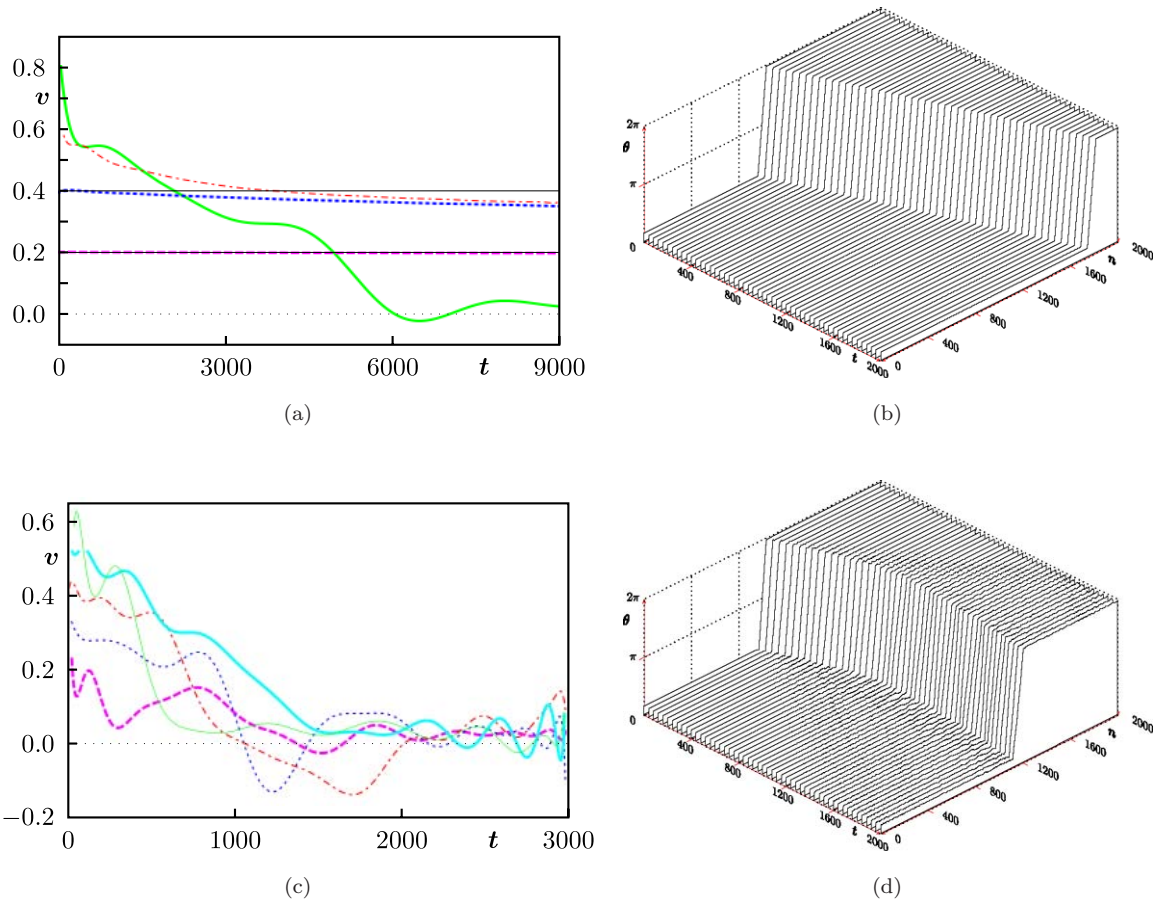


Fig. 4. Time evolution of speeds (in km/s) and profiles of solitons of topological type $(1, 1)$ in the YM model with $m = 7$, $g = 7.5$, $K = 0.02$, $\mu^Y = 11$. *Homogeneous chain:* (a) Evolution of the speed of the center of kinks corresponding to (from bottom to top): $v = 0.2, 0.4, 0.6, 0.71 \simeq c$. The qualitative behavior is identical to the harmonic case. (b) Evolution of the profile of the kink with $v = 0.4$. *HA2 chain:* (c) Evolution of the speed of the center of kinks corresponding to (from bottom to top): $v = 0.2, 0.3, 0.4, 0.5, 0.57 \simeq c$. The motion stops in all cases after between 1000 TU and 2000 TU; for $v \geq 0.4$ in this time the kink has moved by about 4 times its diameter. (d) Evolution of the profile of the kink with $v = 0.4$.

kinks with the same geometric and dynamic parameters used in [40] for topological numbers $(1, 1)$ and velocities $v = 0.4\,\text{km/s}$ and $v = 0.62\,\text{km/s} \simeq c$. In both cases the profiles do not change while phonons are visibly emitted forwards and backwards.

There are two possible explanations for this unexpected behavior: (1) the continuous soliton solutions are unstable, so that any perturbation — no matter how small — would destroy it; (2) the discrete solutions we find do not correspond to any continuous propagating soliton.

There are some facts supporting the second possibility. The profiles of the nontopological angles jump sharply — within a single node — from an angle of about -1 to an angle of about 1 and this remain true when the parameters' values are in a large region of the phase space; a lot of phonons are immediately emitted in both directions even at small speeds, which is exactly what is supposed to happen when the initial profile is not close to a continuous soliton; the nonlinear interactions V_p (pairing) and V_c (the effective confining

potential) increase drastically when the nontopological angles move from their from the equilibrium position, effectively freezing them on some stable static position.

5.2. *Time evolution under Morse pairing potential*

It is well known that using a harmonic potential for the pairing interaction is nothing but a rough approximation: pairing is due to hydrogen bonds, which are highly directional and basically disappear completely once the bases rotate by a few degrees about the respective backbone segment; on the contrary a harmonic potential corresponds to a force which

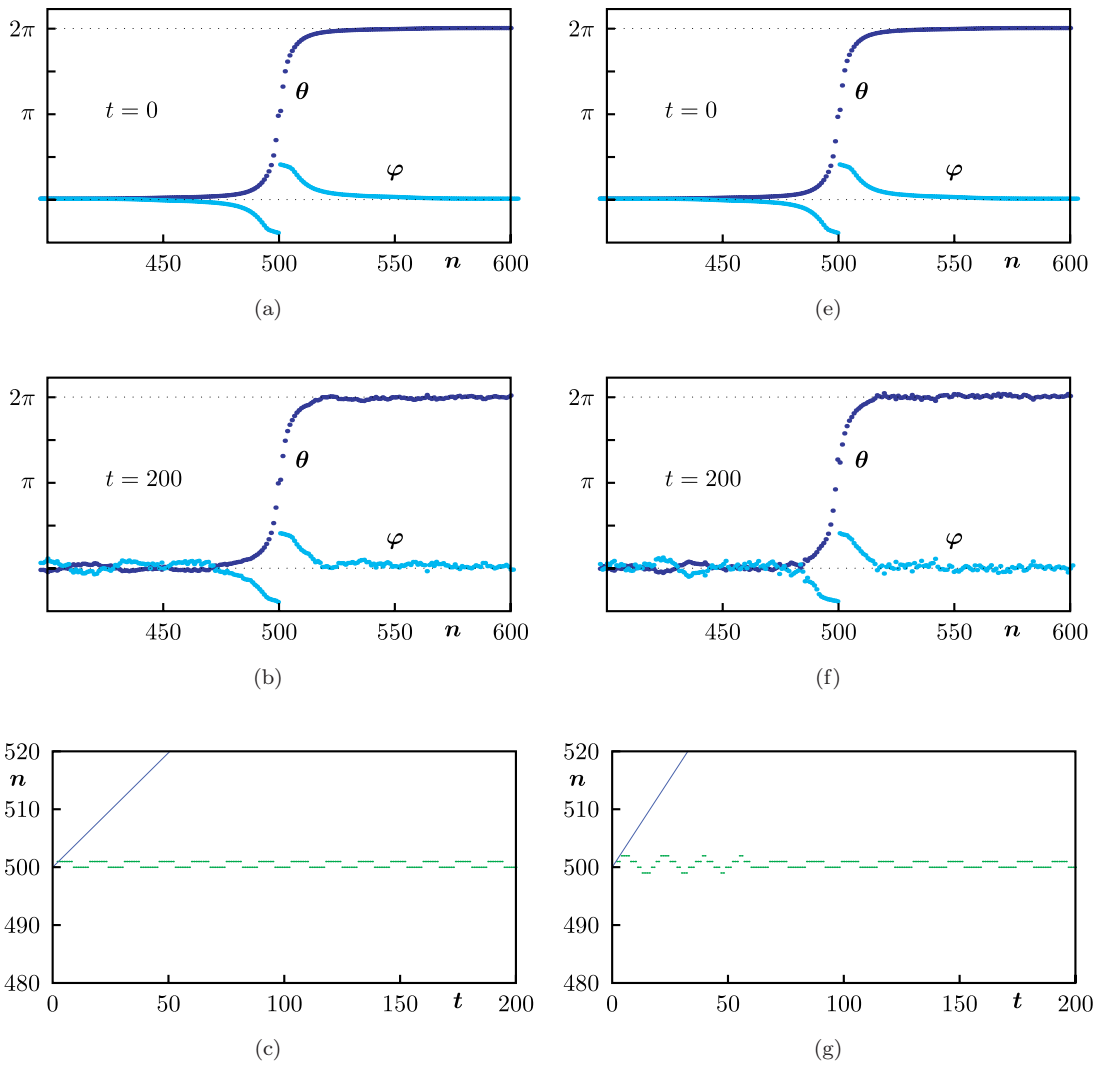
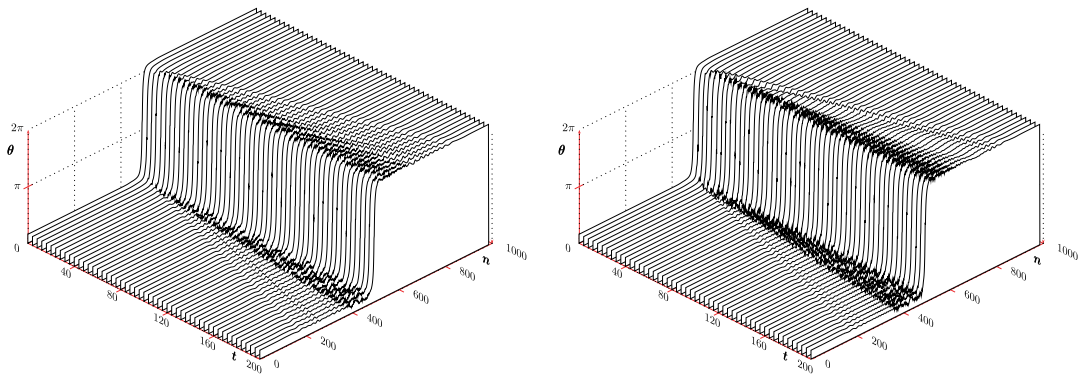


Fig. 5. Profiles and time evolution of kinks in the homogeneous CY model with $i_t = 1.1$, $i_s = 1.3$, $g_t = 0.58$, $g_s = 6.5$, $g_p = 1$, $g_h = 0.03$, and $g_c = 100$ (these are the values used in [40]) and topological numbers $(1, 1)$ for speeds $v = 0.4\,\mathrm{km/s}$ (left) and $v = 0.62\,\mathrm{km/s} \simeq c$ (right). (a,e) Initial profile. (b,f) Evolution of the soliton after 200 TU — the kink did not move but its shape is clearly modified by the continuous emission of phonons in both directions. (c,g) Motion of the soliton center (green thick line) compared to the motion of its continuous counterpart (grey thin line). (d,h) Motion of the soliton — phonon emission is clearly visible in both directions.



(d) (h)

Fig. 5. (*Continued*)

increases linearly with the distance, something that hence is quite different from the physical phenomenon. This is the reason that urged us to introduce a more realistic potential for the pairing interaction and, as it turned out, this is enough to restore the capability of solitons of moving along the chain. In other words, our choice of the Morse potential for the pairing interaction has been naturally suggested by the geometrodynamics of the model itself.

Note that this potential nearby the equilibrium position coincides with the harmonic one for $K_p = 2\kappa a^2$, so it gives rise to the very same dispersion relations. Its main physical improvement with respect to its quadratic approximation is that base-pair forces go exponentially to zero after they get apart by more than a distance of the order of a^{-1}; in particular the energy of Morse solitons is much lower than their corresponding harmonic counterpart. In general it turns out also that Morse profiles are always wider than the corresponding harmonic counterpart (see e.g. Fig. 2(left)) and their limit velocity is higher.

5.2.1. *Yakushevich model*

Motivated by the absence of motion of kinks in our CY model in the harmonic approximation, we turn now to the study of the YM model. In Figs. 4(a) and 4(b), we show profiles and motion of some YM kink. As expected, kinks are much wider than those of the harmonic counterpart. We also observed that the limit speed is about 1.5 times the one of that appearing in the harmonic approximation. For solitons' speed near to the limit speed c, phonon emission is smaller than in the harmonic case, even though this phenomenon might depend crucially on how close is the soliton speed to c. This shows that switching to Morse potential does not alter significantly the behavior of kinks evolution in a homogeneous chain with the Yakushevich geometry (but we will show later that it does alter it for inhomogeneous chains). Conversely, we will show in the following that this switching makes a substantial difference in the case of composite model.

5.2.2. *Composite model*

Now we consider the CYM model. In Fig. 6, we show profiles and motion of kinks with the default values of the coupling constants (see Table 4) and topological numbers $(1, 1)$.

A first observation is that, as already pointed out in [40], the profiles of the nontopological angles are smooth rather than jumpy like it happens in the harmonic approximation (see e.g. Fig. 8 in [40]). This is due to the fact that profiles in the harmonic approximation tend to be too steep. The diameter of the kink at $v = 0.4\,\mathrm{km/s}$ is just slightly wider (69 bp) than

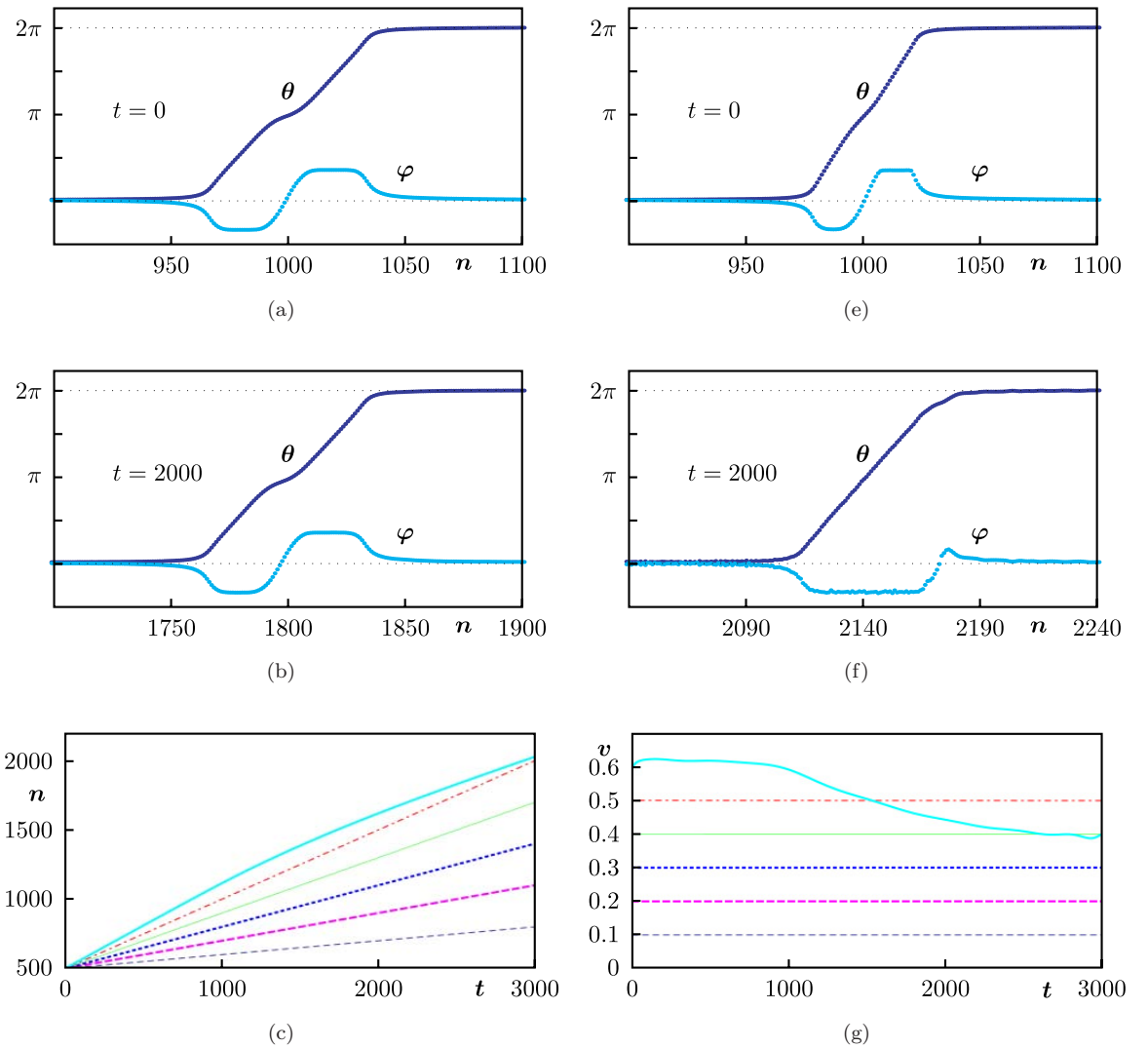
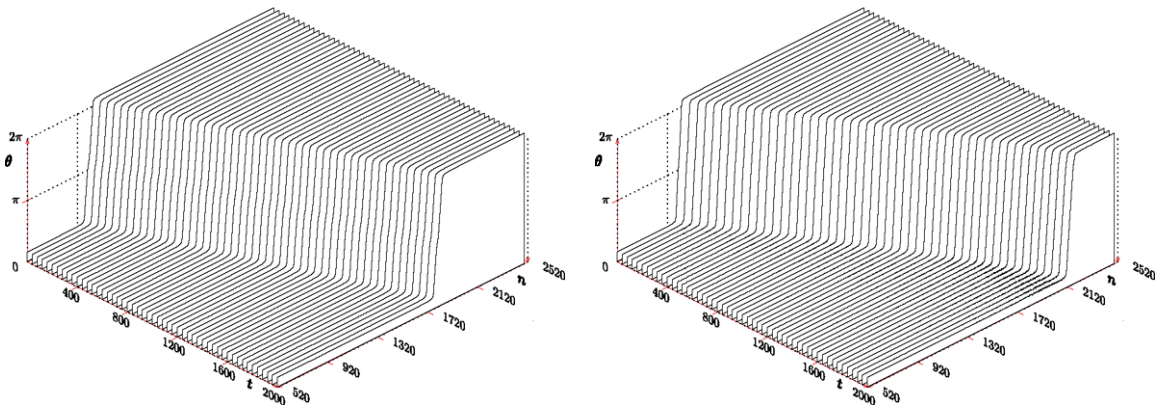


Fig. 6. Profiles and time evolution of kinks with topological numbers $(1, 1)$ in the homogeneous CYM model with $i_t = 1.1$, $i_s = 1.3$, $g_t = 1$, $g_s = 0.3$, $g_h = 0.01$, $g_p = 0.02$, $\mu = 6.3$ and $g_c = 10$. (a) Initial profile for $v = 0.4\,\mathrm{km/s} \ll c$ — the profile is just slightly wider than the corresponding one in the YM model. (b) Evolution of the soliton after 2000 TU — its shape is almost identical to the initial one. (c) Motion of the centers of kinks with speed parameter equal to (from bottom to top) $v = 0.1, 0.2, 0.3, 0.4, 0.5, 0.62 \simeq c$. Except for the case $v = 0.62$, very close to the limit speed, all motions have speed very close to constant. (d) Profiles of the soliton in the first 2000 TU — no phonons emission is visible. (e) Initial profile for $v = 0.62\,\mathrm{km/s} \simeq c$ — diameter is just slightly smaller than the $v = 0.4\,\mathrm{km/s}$ one. (f) After 2000 TU the diameter of the topological component is just slightly wider and its shape slightly changed. The diameter of the non topological component also changed just slightly but its shape changed a lot: the minimum got wider while the maximum almost disappeared. (g) Evolution of the speed of the center of kinks of point (c). (h) Profiles of the soliton in the first 2000 TU — phonons are clearly emitted in both directions but much more in the opposite direction.



(d) (h)

Fig. 6. (*Continued*)

the corresponding one in the YM model (64 bp) and the limit speed $c \simeq 0.62$ is just slightly smaller than the corresponding one in YM ($c \simeq 0.71$). The motion appears to be very close to the continuous counterpart for $v = 0.4\,\mathrm{km/s}$ and slightly decelerated after a while when closer to the limit speed. Nevertheless, like in the YM model, the emission of phonons looks absent in the first case and not very high in the second.

We conclude the section by pointing out again the substantial qualitative difference, in the CY model, between dynamics with the Morse and harmonic pairing potential. The CY and Y model behave qualitatively in the same way under a Morse pairing potential. Conversely, their behavior under an harmonic pairing potential is drastically different: in the CY model kinks stop immediately their motion while in the Y model they move happily like in the YM case. We believe this is due to the fact that the CY model has a geometry detailed enough to feel the roughness of the harmonic approximation and this disrupts the delicate equilibria that make possible the motion of solitons along the chain. It is a remarkable fact that a refinement of geometry of the model requires, in turn, a refinement of its dynamics.

6. Soliton's Motion in Inhomogeneous DNA Models

In this section we discuss our results on the dynamics of kinks in inhomogeneous DNA models, in particular for a real base sequence (the Human Adenovirus 2). To model the inhomogeneities in the real DNA sequence we now use the exact Hamiltonian described in Sec. 2, where differences in both the dynamical parameters of the bases (masses, momenta of inertia) and in the strength of the hydrogen bonds for the base-pairs A-T, G-C are taken into account.

It is well known that propagation of solitons in non-homogeneous molecular chains is a rather complicate issue. One expects the motion of the soliton to depend crucially both on the interaction and on the spatial distribution of the inhomogeneities in the chain. Solitons behave as point-particles so that inhomogeneities may act as an effective potential affecting their motion. For instance, they may be trapped in some potential well and keep oscillating inside it or they may bounce back when scattered by a potential barrier. Generically, they

will emit phonons while moving, resulting in a deceleration that will ultimately bring the soliton to stop. Moreover the soliton parameters, e.g. its energy, diameter and maximum distance reached, vary quite a lot depending on the particular distribution of inhomogeneities along the chain and on the starting point of the soliton.

Therefore we are not so much interested in the motion of a single soliton but rather in the "average" properties of solitonic propagation all along the DNA chain. In particular our aim is investigating about the existence, within the range of our geometric and dynamical parameters and all along a DNA chain, of kink-like torsional configurations whose diameter is reasonably narrow (about 60 bp) and which are able to move far enough to be relevant for the DNA transcription process, that is over at least 2–300 base-pairs.

In the rest of the section we will use as base sequence for our simulations the one found in the DNA of the Human Adenovirus 2[i] (HA2), whose length is 35937 bp; we will also compare some of our results with those obtained for random chains of the same length.

Similarly to what we have done in Sec. 5, we discuss separately the time evolution of solitons under harmonic potential (Subsec. 6.1) and under a Morse potential (Subsec. 6.2), respectively for the Y model (Subsecs. 6.1.1 and 6.2.1) and for the CY model (Subsecs. 6.1.2 and 6.2.2).

6.1. *Time evolution under harmonic pairing potential*

6.1.1. *Yakushevich model*

In Figs. 3(c) and 3(d), we show profiles and time evolution of kinks with topological number $(1, 1)$ for the inhomogeneous Y model in the HA2 chain for speeds $v = 0.2\,\mathrm{km/s}$ and $v = 0.38\,\mathrm{km/s} \simeq c$. In the first case the kink tries to start its motion but it is stopped by a potential barrier and starts bouncing back and forth emitting phonons until its kinetic energy is over. In the second case the energy of the kink is enough to make it pass the barrier but it bounces back after about 40 bp, not even its own diameter, and starts oscillating, trapped within some potential well.

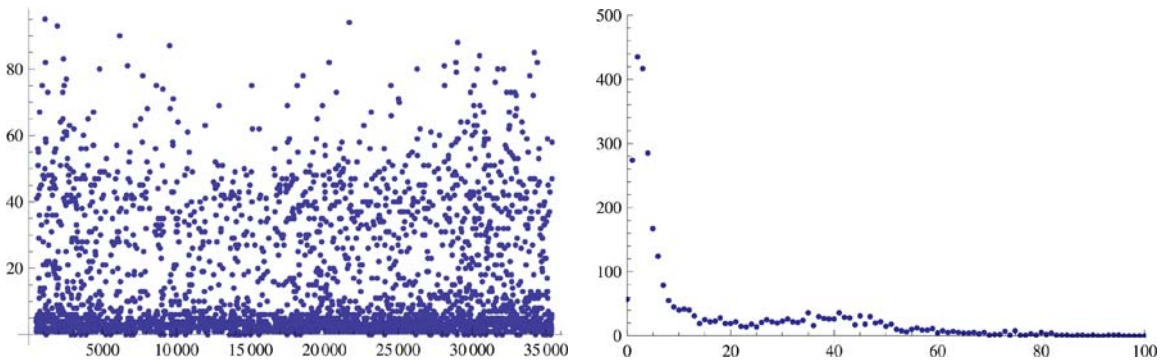
In Fig. 7, we show the distance ran by kinks as a function of their starting position on the HA2 chain and the corresponding statistical distribution. As expected, the distribution of lengths ran by the kinks is a Gaussian centered about a distance of a few bp. In concrete, while a few kinks manage to move up to 100 bp, 90% of them move by less then 41 bp.

This suggests that solitons in the inhomogeneous version of the Y model cannot be used to describe phenomena like transcription or duplication of DNA because kinks are not able to move for large enough distances. In the case of a (physically more realistic) Morse-type pairing potential this problem disappears.

6.1.2. *Composite model*

Inhomogeneities in the chain clearly make motion harder rather than easier. Since kinks were not able to move in the homogeneous CY model, we expected no motion in the inhomogeneous CY model too and numerical experiments fully confirmed this.

[i]http://www.binf.gmu.edu/adenovirus_database/images/h2.htm.



(a) (b)

Fig. 7. (a) Maximum distance reached vs. initial center for $(1,1)$ kinks in the inhomogeneous Y model in a HA2 chain with constants $m = 7$, $g = 21$, $K = 1$ and speed $v = 0.38\,\mathrm{km/s}$; (b) Number of kinks in the previous plot vs. maximum distance reached: 50% of the kinks run by no more than 5 bp, 90% of them by no more than 41 bp.

6.2. *Time evolution under Morse pairing potential*

6.2.1. *Yakushevich model*

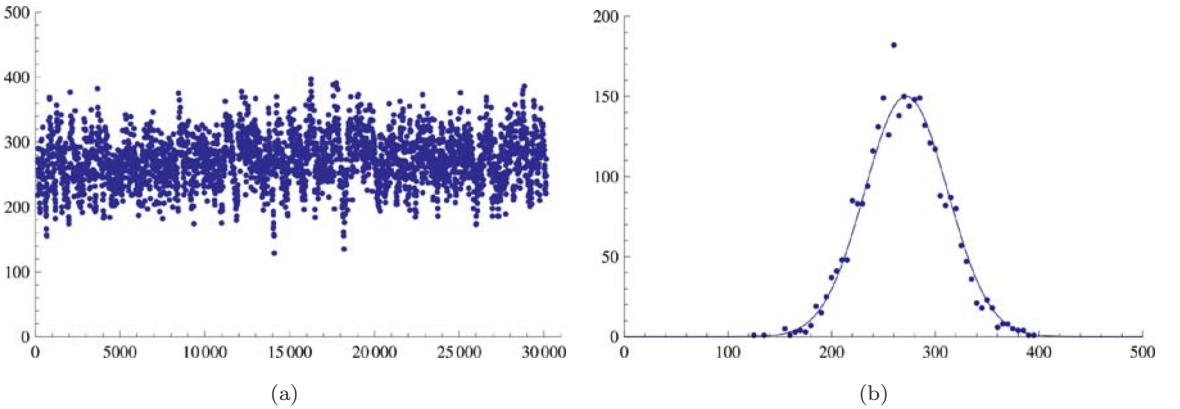
The motion of kinks along the chain shows different features depending on the region the soliton starts from. In Figs. 4(c) and 4(d), we show profiles and motion of a kink in the inhomogeneous YM model in the HA2 chain sequence and with $v = 0.4\,\mathrm{km/s}$ and one with $v = 0.57\,\mathrm{km/s}$. The profiles are slightly more "wavy" than in the homogeneous case and emit phonons, as expected, even far from the limit velocity; moreover phonons in this case are clearly emitted in both directions and, again, much more when the velocity is close to the limit speed. The soliton evolution is close to the continuous limit in the first few hundreds TU after that the soliton either starts making small oscillations about some point or bounces back and eventually starts oscillating within some potential well.

Every kink along the chain has a behavior of this kind and both diameter and energy do not change much (within 10^{-2}) from point to point; what does change a lot is the maximum distance the soliton reaches, as shown in Fig. 8. Soliton's diameters are about 56 bp. The kink travels for considerable distances before stopping owing to phonon emission. The average length run by kinks is about 250 bp, i.e. about 4 times the diameter of the soliton. The distance reached by the soliton is always between 100 bp and 400 bp, namely between 2 and 8 times their diameter.

6.2.2. *Composite model*

As in the YM case, even in the CYM model the distances run by kinks in inhomogeneous chains change a lot from point to point.

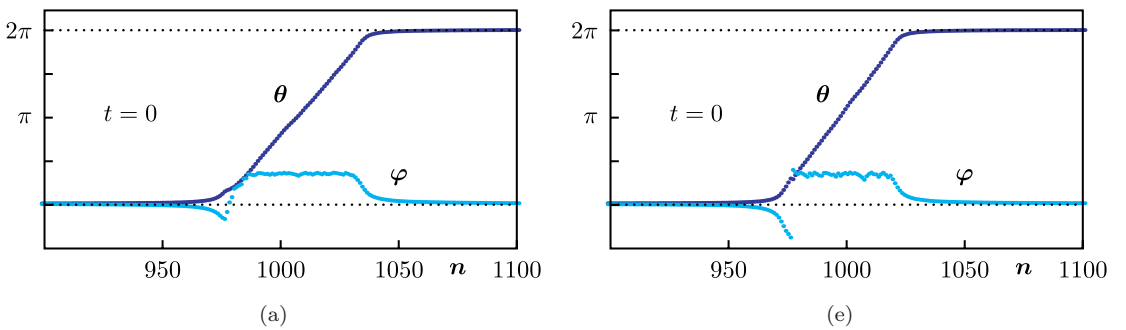
In Fig. 9, we show profiles and motion of a kink for the CYM model in a HA2 chain sequence and with topological numbers $(1,1)$ and $v = 0.4\,\mathrm{km/s}$ and one with $v = 0.45\,\mathrm{km/s} \simeq c$. The shapes of profiles are just slightly different from the homogeneous case, especially the nontopological ones, and the motion of center is qualitatively the same as in YM. Again, phonons are emitted in both directions and much more when v is



(a) (b)

Fig. 8. (a) Maximum distance reached vs. initial center for $(1,1)$ kinks in the inhomogeneous YM model in a HA2 chain with constants $m = 7$, $g = 7.5$, $K = 0.02$, $\mu^Y = 11$ and speed $v = 0.4\,\mathrm{km/s}$; (b) Number of kinks in the previous plot vs. maximum distance reached: 90% of the kinks run between 197 bp and 347 bp. Since the diameter of such solitons is about 60 bp, these kinks move between 3 and 6 times their diameter.

close to the limit velocity. Note also that phonons seem to propagate at constant speed even though the medium is nonhomogeneous. The first kink moves for about 500 bp at almost constant speed before starting to decelerate and is able to move for about 400 bp; the second starts decelerating almost immediately and stops after less than 200 bp. This corresponds



(a) (e)

Fig. 9. Profiles and time evolution of kinks with topological numbers $(1,1)$ in the inhomogeneous CYM model in a HA2 chain and constants $i_t = 1.1$, $i_s = 1.3$, $g_t = 1$, $g_s = 0.3$, $g_h = 0.01$, $g_p = 0.02$, $\mu = 6.3$ and $g_c = 10$ at speed $v = 0.4\,\mathrm{km/s} < c$ (left) and $v = 0.45\,\mathrm{km/s} \simeq c$ (right). (a) Initial profile for $v = 0.4\,\mathrm{km/s}$ — the profile is just slightly wider than the corresponding one in the Y model. (b) Evolution of the soliton after 2000 TU — its shape is almost identical to the initial one but a little "wavy", as expected. (c) Motion of the centers of kinks with center at $n = 1000$ and speed parameter equal to $v = 0.1$ (thin, dashed), 0.2 (magenta, thick, dashed), 0.3 (blue, thick, dotted), 0.4 (green, thin), $0.45 \simeq c$ (red, dotted and dashed). The kink with $v = 0.1$ loses immediately its kinetic energy, basically does not move at all; all others move for at least 100 bp before stopping and their maximum distance ran on the chain increases with the velocity parameter v except for the case $v = 0.45$, which is very close to the limit speed. (d) Profiles of the soliton in the first 2000 TU — very little phonons emission is visible in both directions. (e) Initial profile for $v = 0.45\,\mathrm{km/s}$ — diameter is just slightly smaller than the $v = 0.4\,\mathrm{km/s}$ one but the shape is significantly more "shaky". (f) After 2000 TU the diameter of the topological component is just slightly wider and its shape slightly changed but the presence of phonons is much more evident. (g) Evolution of the speed of the center of kinks of point (c). (h) Profiles of the soliton in the first 2000 TU — phonons are clearly emitted in both directions but much more in the opposite direction.

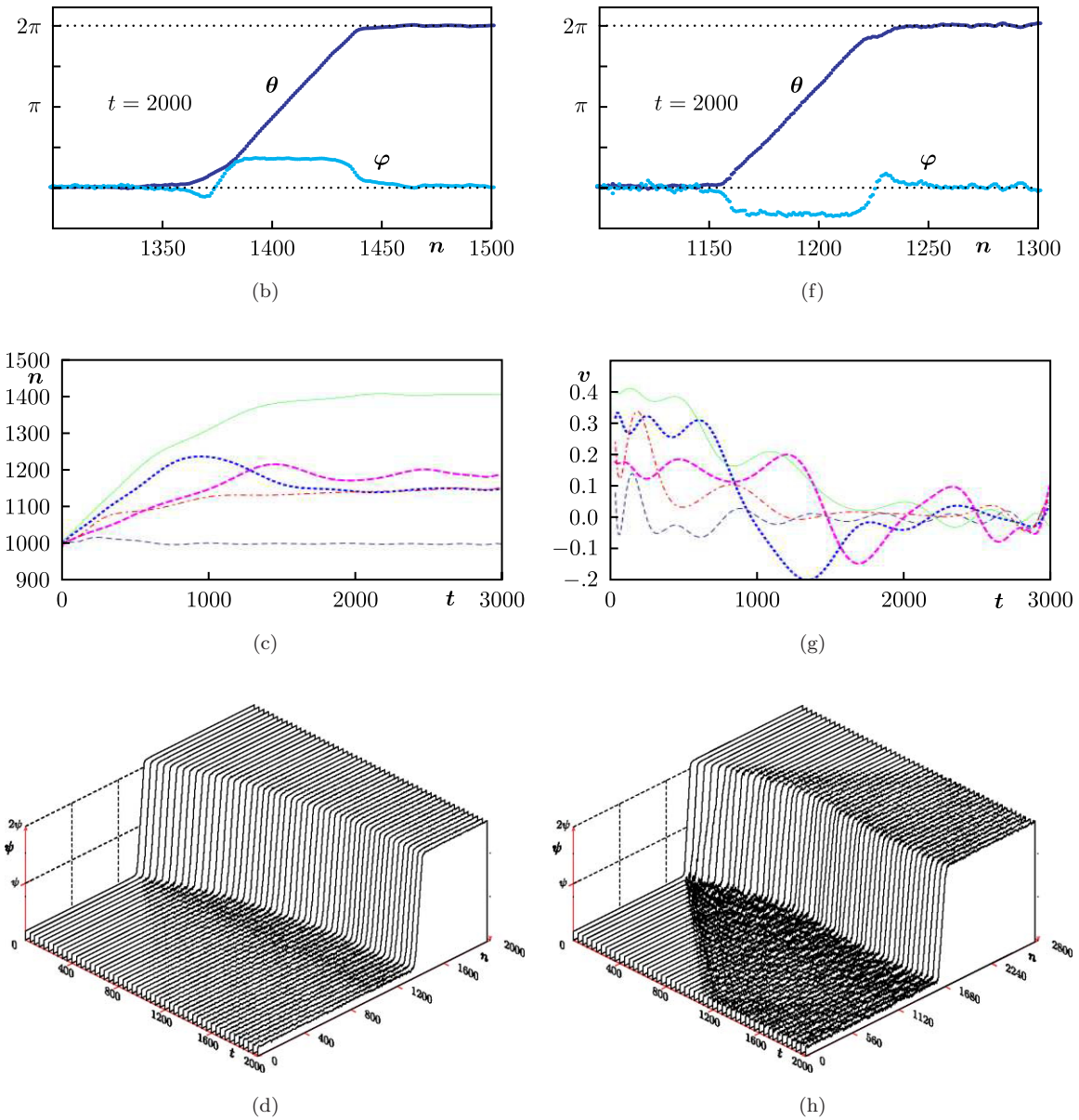


Fig. 9. (*Continued*)

to the fact that, close to the limit speed, there is a big emission of phonons and therefore the kink decelerate more rapidly.

The most important result of this paper is illustrated in Fig. 10. In the left hand side of the figure we show both the distances reached by $(1, 1)$ kinks as function of the starting point on the HA2 chain, the distribution of the maximum distance reached by the kinks and the mean distance run by $(1, 1), (1, 0), (0, 1)$ solitons versus their speed. The diameter of the kinks is about 60 bp. The maximum distance run by kinks lie all between 150 bp and 650 bp, i.e. about 2-10 times their diameter. The mean distance run is 353 bp, that is about 6 times the diameter of the soliton; and 90% of kinks run between 237 bp and 468 bp,

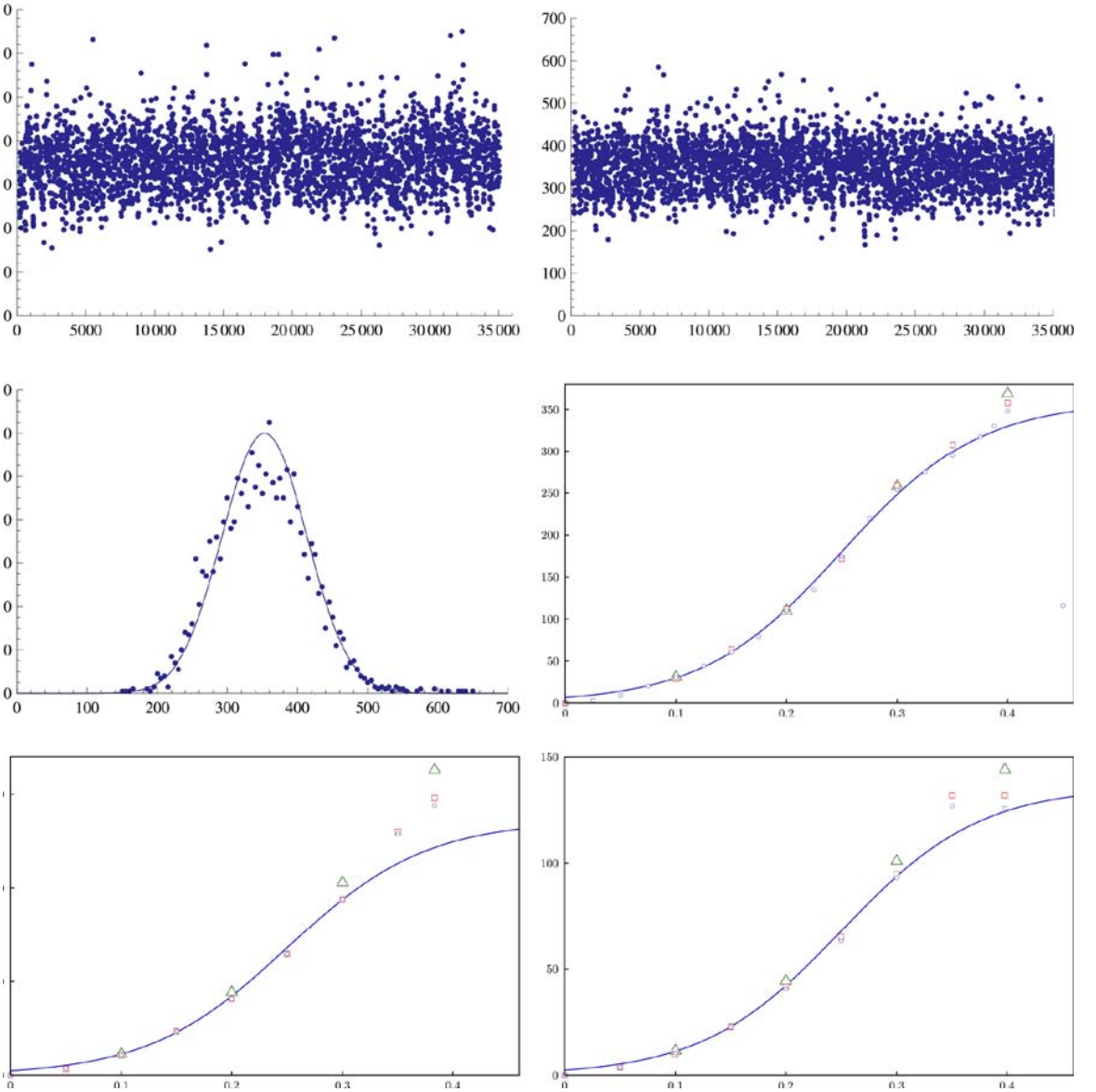


Fig. 10. (First row) Max distance reached vs. initial center for $(1,1)$ kinks in the CYM model in a HA2 chain (left) and in a random chain (right) with constants $I_t = 1.1$, $I_s = 1.3$, $g_t = 1$, $g_s = 0.3$, $g_p = 0.02$, $\mu = 6.3$, $g_h = 0.01$, $g_c = 10$ and $v = 0.4\,\mathrm{km/s}$. (Second row, left) Number of kinks in the HA2 chain vs. maximum distance reached: the mean distance run is 353 bp and 90% of kinks run between 237 bp and 468 bp. Since the diameter of such kinks is about 60 bp, this means that they run between 4 and 8 times their diameter. (Second row, right) Mean distance run vs. speed for $(1,1)$ kinks with coupling constants as above in the interval between 0 speed (static solitons) and the limit speed $c \simeq 0.45$ for a HA2 chain (blue, circles), a random chain (red, squares) and a chain of human mitochondrial DNA (green, triangles). The data plot, outside of some small neighborhood of the extremes of the interval $[0, 0.45]$, is well fitted by the graph of the function $y = 180\tanh(8x - 2) + 180$. Note that when the speed of the kink gets too close to the limit speed the distance ran gets very small. (Third row) Analogue plots for kinks with topological number $(1,0)$ (left) and $(0,1)$ (right). In this case the data plots are both well approximated by the graph of the function $y = 68\tanh(8x - 2) + 68$ and in both cases deviations are more pronounced close to the limit speed $c \simeq 0.4$.

i.e. between 4 and 8 times their diameter. The picture is similar to the one corresponding to the Y model but is shifted above by about 50%, i.e. kinks in the CYM model move in average 50% more than the corresponding kinks in the YM model even though they have similar energies and diameters.

A second important result is illustrated on the upper right hand part of Fig. 10, where we show that the average length traveled by the solitons and the soliton distribution does not depend on the particular combinatorics of the DNA sequence. Indeed we evaluated the same quantities on a random sequence of same bp length and found the very same statistics (Fig. 10(right)).

This means that the DNA sequence, which is of course fundamental in all biological processes, does not play a significant role in torsional dynamics, i.e. from the torsional dynamic point of view all "generic" chains look more or less the same. Note that this fact means that the mechanical, torsional, features of DNA which allow for the existence of twist solitons are "almost" decoupled from the combinatorics aspects of DNA, which encode genetic information.

7. Effects of Dissipation and External Forces on Soliton's Motion

Until now we have modelled the DNA chain as a conservative system and disregard all the interaction of the DNA molecule with its environment. However, DNA is an overdamped system subject to forces, including random ones, arising from interactions with its environment and, obviously, these interactions are relevant for understanding functional properties of DNA such as the transcription process. In this section we will discuss these effects for the case of the inhomogeneous Y model both with a harmonic and a Morse pairing potential; the discussion of kink propagation in presence of damping effects and external forces in the CY model is deferred to a later work for the sake of brevity. Our result represent the generalization to the inhomogeneous chain of previous results obtained by Yakushevich for the homogeneous chain [14].

As a first and rough approximation we can, following Yakushevich [14], describe the interaction of the DNA with its environment by introducing, in the equations of motion for θ in the Y model of Subsec. 2.1, a (torsional) damping term and an external (torsional, constant) force:

$$(-\nu_0\dot{\theta}_{n,i} + F_0)E_0. \tag{7.1}$$

Notice that we are analyzing rotational dynamics so that Eq. (7.1) describes a torque acting on the system, which has the dimension of an energy and can be measured either in kJ/mol or Nm/mol. The parameter F_0 is dimensionless while ν_0 has the dimension of a time.

Since we are working with a discrete DNA chain of finite length (between 10^3 and 10^4 sites depending on the parameters), the presence of a constant external force at both ends of the chain is not compatible with the boundary conditions, which fix $\theta = 2n\pi$. To solve this problem, F_0 is set to zero at the 200 sites at each end of the chain.

The range of variation for the parameters ν_0 and F_0 can be found in the literature (see [14] and the references therein). In our computations we used for the damping coefficient the model value $\nu_0 = 6 \cdot 10^{-2} TU$, so that $\nu_0 v_0/\delta = 6 \cdot 10^{-2}$. For F_0 we have used values ranging from $5 \cdot 10^{-1} \sim 110\,\mathrm{kJ/mol}/E_0$ to $10^{-2} \sim 2.2\,\mathrm{kJ/mol}/E_0$.

We have determined, numerically, the time evolution of the kinks with topological numbers $(1, 1)$ when a dissipative and/or an external force are present for: (1) the homogeneous Y model (harmonic pairing potential) and YM model (Morse pairing potential); (2) the inhomogeneous Y and YM models in a HA2 chain sequence. The results of our computations are summarized in Figs. 11 and 12.

For what concerns time evolution of the kink in homogeneous chains, our results fully confirm the previous results by Yakushevich [14] (see Figs. 11(a) and 11(b)) and extend them to the YM model (see Figs. 12(a) and 12(b)). There exists a critical value F_{0c} of the external force above which the kink can propagate at constant speed also in the presence of dissipation. $F_{0c} = 10^{-4}$ in the case of the Y model, whereas $F_{0c} = 10^{-5}$ for the YM model. There is also an upper critical value F_{0C} for the external force, above which the soliton is

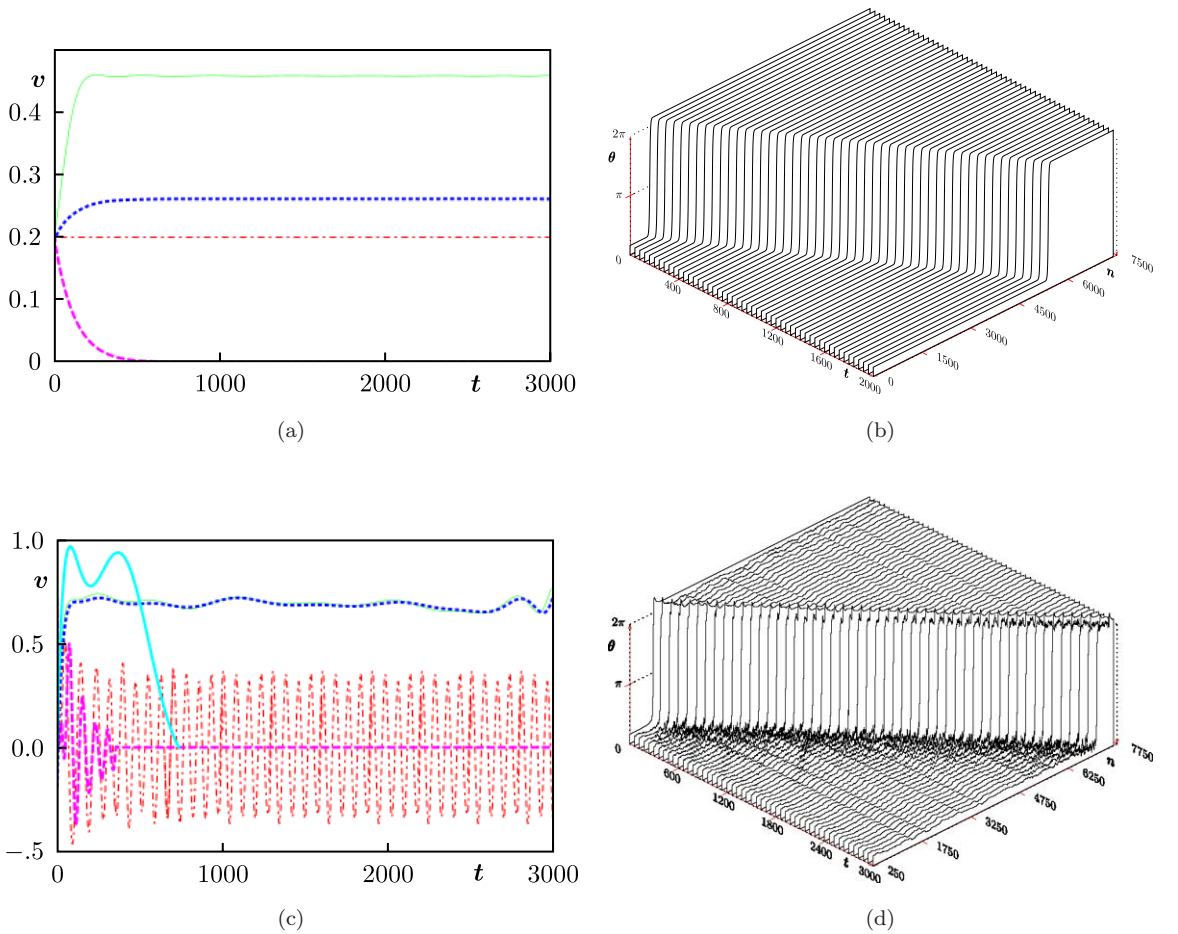


Fig. 11. Time evolution of speeds (in km/s) and profiles of solitons of topological type $(1, 1)$ in case of a DNA chain in presence of dissipation ν and constant external force F. Our model value for the dissipation coefficient is $\nu_M = 6 \cdot 10^{-2}$. **Homogeneous Y model**, $v = 0.2$, $m = 7$, $g = 21$, $K = 1$; (a) Speed vs Time with (ν_0, F_0) equal to $(0, 0)$ (red, dot-dashed), $(\nu_M, 0)$ (thick, pink, dashed), $(0, 10^{-2})$ (green), $(\nu_M, 10^{-2})$ (blue, dashed); (b) Motion of the kink with $(\nu_0, F_0) = (\nu_M, 10^{-2})$. **Inhomogeneous Y model (HA2 sequence)** (c) Speed vs Time with (ν_0, F_0) equal to $(0, 0)$ (red, dot-dashed), $(\nu_M, 0)$ (thick, pink, dashed), $(0, 10^{-1})$ (green), $(\nu_M, 10^{-1})$ (blue, dashed), $(\nu_M, 5 \cdot 10^{-1})$ (cyan, thick). (d) Motion of the kink with $(\nu_0, F_0) = (\nu_M, 10^{-1})$.

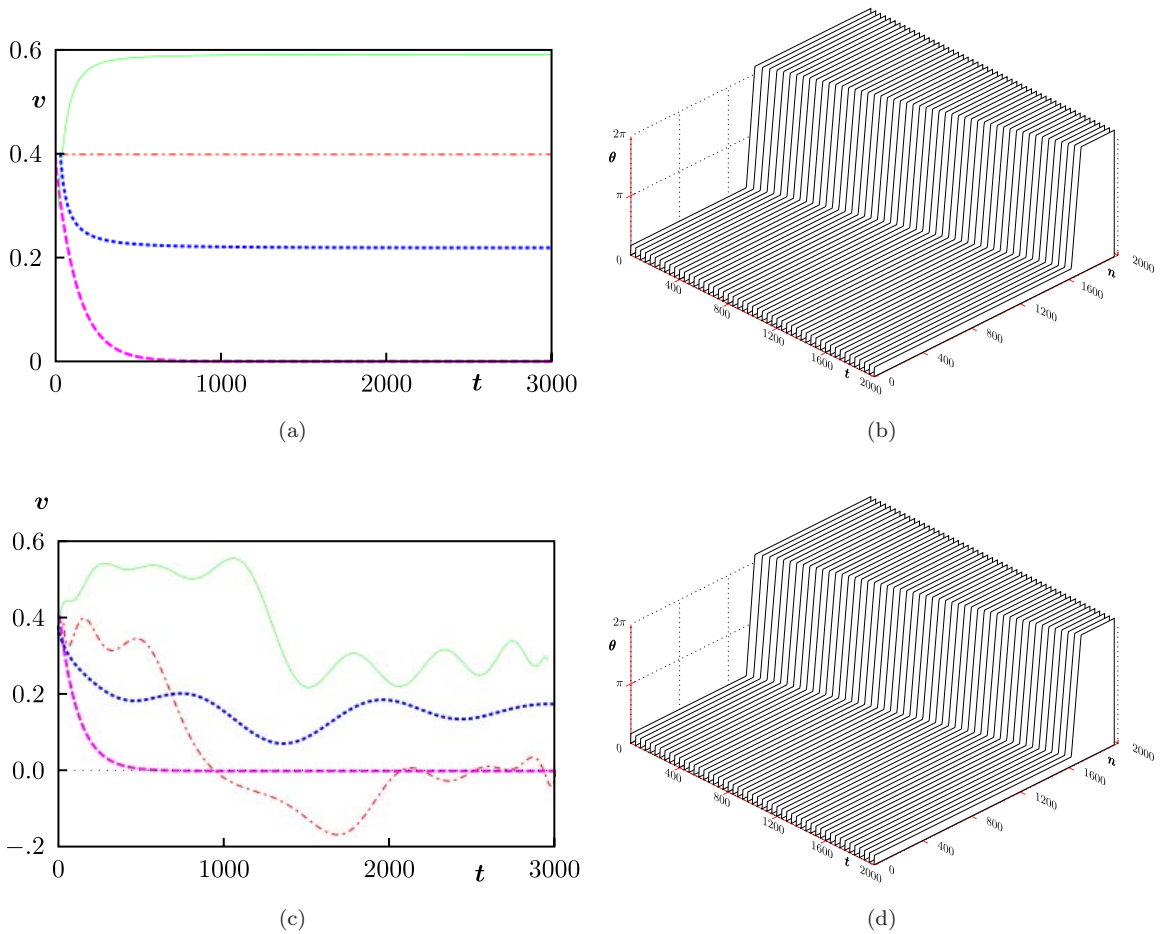


Fig. 12. Time evolution of speeds (in km/s) and profiles of kinks of topological type $(1,1)$ in a DNA chain in presence of dissipation ν and constant external force F_0. Our model value for the dissipation coefficient is $\nu_M = 6 \cdot 10^{-2}$. **Homogeneous YM model**, $v = 0.4$, $m = 7$, $g = 7.5$, $K = 0.02$, $\mu^Y = 11$; (a) Speed vs Time with (ν_0, F_0) equal to $(0,0)$ (red, dot-dashed), $(\nu_M, 0)$ (thick, pink, dashed), $(0, 10^{-3})$ (green), $(\nu_M, 10^{-3})$ (blue, dashed); (b) Motion of the kink with $(\nu_0, F_0) = (\nu_M, 10^{-3})$. **Inhomogeneous YM model (HA2 sequence)**, $v = 0.4$, $m = 7$, $g = 7.5$, $K = 0.02$, $\mu^Y = 11$; (c) Speed vs Time with (ν_0, F_0) equal to $(0,0)$ (red, dot-dashed), $(\nu_M, 0)$ (thick, pink, dashed), $(0, 10^{-3})$ (green), $(\nu_M, 10^{-3})$ (blue, dashed). (d) Motion of the kink with $(\nu_0, F_0) = (\nu_M, 10^{-3})$.

strongly distorted and eventually becomes too steep and stops moving or the DNA chain as a whole opens and closes ($F_{0C} = 1$ for the Y model and $F_{0C} = 2.5 \cdot 10^{-3}$ for the YM model).

The most interesting results concern the time evolution of the kink in the inhomogeneous YM model with a HA2 chain. When an external force above the critical value $F_{0c} = 7.5 \cdot 10^{-4}$ is present, the kink moves all the time with nonvanishing speed both in absence of dissipation (green line in Fig. 12(c)) and with dissipation (blue, dashed line in Fig. 12(c)). Because now we are dealing with an inhomogeneous chain, the soliton speed does not approach asymptotically to a constant value (as in the homogeneous case) but oscillates around a nonvanishing value. Also in this case we have an upper critical value for the external force given by $F_{0C} = 2.5 \cdot 10^{-2}$.

In the case of the inhomogeneous Y model with a HA2 chain (see Fig. 11(c)) we have similar results but the window of the external force for which kinks propagates along the chin without stopping is now very narrow: $7.5 \cdot 10^{-2} \leq F_{0c} \leq 2.5 \cdot 10^{-1}$. This peculiar behavior of the inhomogeneous Y model can be probably traced back to the features of the harmonic pairing potential that make the kink more sensible to inhomogeneities on the chain.

8. Motion of Kinks at Low Speeds

So far we considered only kinks with speeds of the same order of magnitude of DNA's speed of sound, which is of the same order of magnitude of the limiting speed c of solitons. The velocity of transcription and replication though is much smaller, between 50 bp/s and 10^3 bp/s [52], namely at least about 10 orders of magnitude smaller than our reference speed v_0. Since we are bound (see Subsec. 4.2) to use a time unit about 10^{-13} s, the computing time needed to study numerically such a slow motion is way too long for a concrete simulation.

A second difficulty, which prevents us to investigate numerically soliton motion with speeds of order of magnitude 10^3 bp/s, is the presence of the well-known Peyerls–Nabarro potential (e.g. see [68]). This is a discreteness effect which implies that, in the homogeneous case, there is a lower bound on kinks' speeds equal to $6 \cdot 10^{-3}$ in Y and to $2 \cdot 10^{-4}$ in YM. In view of these difficulties in this section we investigate soliton motion at speeds which are compatible with the presence of the Peyerls–Nabarro potential and with our tiny time unit. These speeds are up to 4–5 orders of magnitude smaller than v_0. For simplicity we will also only consider the case of $(1, 1)$ kinks in the Y and YM models.

Considering soliton speeds of order of magnitude $10^{-5}v_0$ we are still far away by 5 order of magnitude from the speed of transcription and replication in real DNA. However we should keep in mind that in these processes are actually involved many external factors, not last the RNA polymerase, which acts as external forces on the DNA chain. The question therefore is whether the Y and YM models support kinks moving with speeds slower than the lower bounds for the free solitons *after* external forces are put in play. For this reason we we discuss here the case of DNA subject to a constant external force F and a friction ν.

Because at this stage our investigation is purely qualitative we choose arbitrarily the values of F_0 and ν_0 (see the previous section). We use the inhomogeneous sequence of HA2 in our simulations. The results of our numerical computations are shown in Fig. 13. They show clearly that in the presence of external forces and dissipation soliton motion at roughly constant speed is possible at speeds of order $10^{-5}v_0$. For example, by choosing $F_0 = 0.4$ and $\nu_0 = 5 \cdot 10^3 TU$ in Y the kink moves at a speed of about $2 \cdot 10^{-4}$, about a order of magnitude smaller than the allowed speed for free kinks, and by choosing $F_0 = 10^{-3}$ and $\nu_0 = 256TU$ in YM the kink moves at speed of about $4 \cdot 10^{-5}$, again about a order of magnitude slower than the allowed speed for free kinks. Numerical experiments seem to indicate that, in both models, by increasing the friction the speed of the kink slows down continuously.

This analysis suggests that, even though no kink can freely move in Y and YM at the real speed for transcription and replication, such speeds can be reached under the action of an external force and therefore this mechanism could play a role in these phenomena.

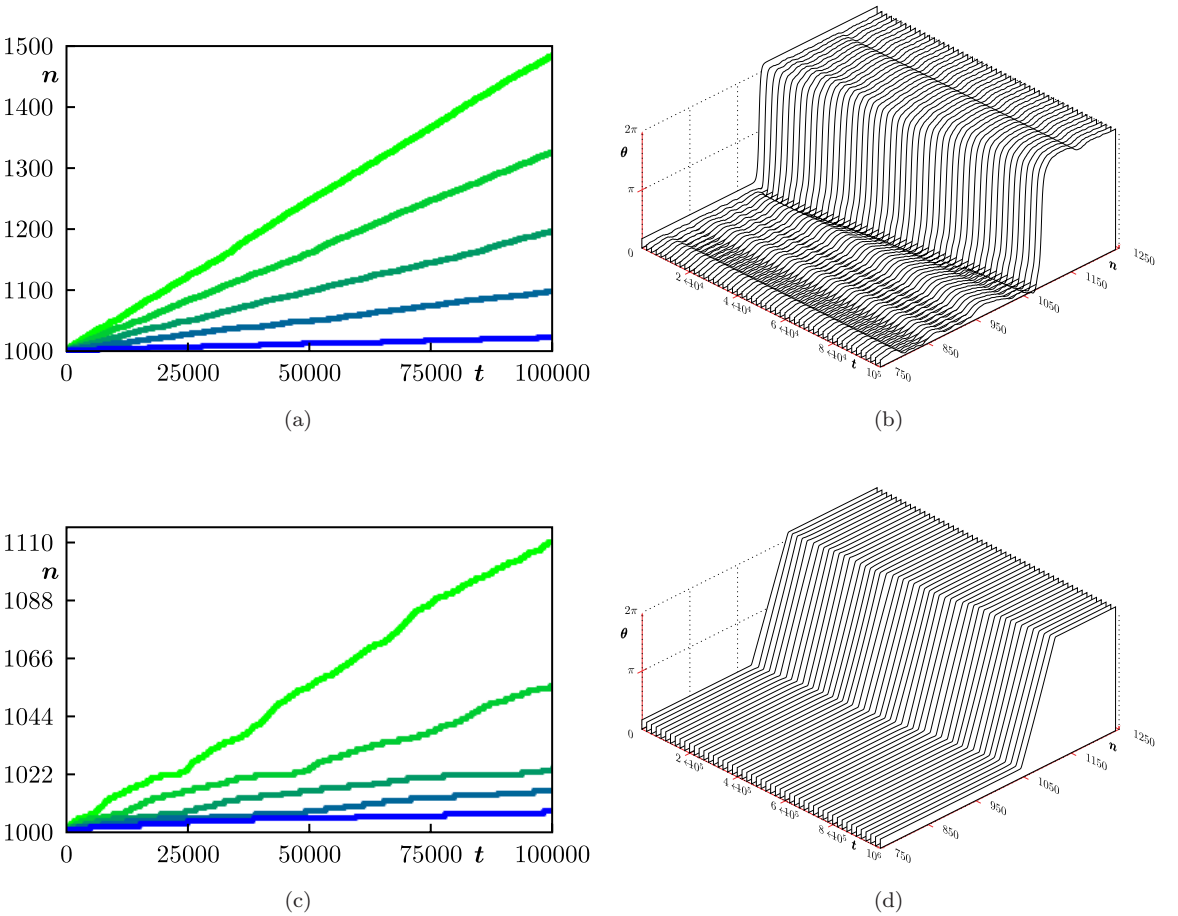


Fig. 13. Time evolution of centers and profiles of kinks of topological type $(1,1)$ in a DNA chain in presence of dissipation ν and constant external force F at low speed. **Y model**, $m = 7$, $g = 21$, $K = 1$, $F_0 = 0.4$; (a) Evolution of the center of kinks corresponding to the dissipation coefficients (from top to bottom) $\nu_0 = 200, 300, 500, 1000, 5000$. The corresponding averaged speed are $v = 4.7 \cdot 10^{-3}, 3.1 \cdot 10^{-3}, 1.9 \cdot 10^{-3}, 9.3 \cdot 10^{-4}, 2.1 \cdot 10^{-4}$. (b) Evolution of the profile of the kink with $\nu = 1000$. There is clearly a large emission of phonons but the kink's profile moves at reasonably constant speed without changing sensibly. **YM model**, $m = 7$, $g = 21$, $K = 1$, $F_0 = 0.4$; (c) Evolution of the center of kinks corresponding to the dissipation coefficients (from top to bottom) $\nu_0 = 8, 16, 32, 64, 128$. The corresponding averaged speed are $v = 1.1 \cdot 10^{-3}, 5.5 \cdot 10^{-4}, 2.4 \cdot 10^{-4}, 1.5 \cdot 10^{-4}, 7.1 \cdot 10^{-5}$. (d) Evolution of the profile of the kink with $\nu = 128$. There is no visible emission of phonons and the kink's profile moves at reasonably constant speed without changing sensibly.

9. Conclusions

In this paper we have performed a detailed and comparative numerical investigation of the initial profiles and time evolution of solitons in models describing DNA nonlinear torsional dynamics. In our investigations we have considered the most important mesoscopic models for DNA torsional dynamics with parameters entering in the model within the physical range allowed by the experimental results. We have studied the propagation of solitons in models with different geometries (the Y and the CY model), different pairing potentials (harmonic and Morse), and different uniformity features of the DNA chain (homogeneous and inhomogeneous chains). The inhomogeneities in a real DNA sequence have been modelled by

considering the differences in both the dynamical parameters of the bases (masses, momenta of inertia) and in the strength of the hydrogen bonds for the base-pairs A-T, G-C. Here is a summary of our main results.

- Although the existence of static kinks is a generic and robust feature of basically all continuous and discrete models of DNA nonlinear torsional dynamics, for discrete models time evolution of the soliton is rather selective. The time evolution of the soliton shows high sensitivity to the interplay between geometrical, dynamical and uniformity features of the DNA model.
- The simplest model from the geometrical point of view — the Y model — is less sensitive to the choice of the pairing potential for the base-pairs. In the case of the Y model, solitons propagate along the chain both using a harmonic and a Morse pairing potential. Conversely, in the case of the CY model propagation of solitons is possible only when the pairing interaction is modelled by a Morse potential. This gives a nice coupling between geometry and dynamics: a more realistic modelling geometry of the DNA chain necessarily requires a more realistic modelling for the pairing interaction.
- Independently of the geometrical and uniformity features of the DNA chain, soliton propagation discriminates between the Morse and the harmonic potentials. In all our simulations (Y and CY model, homogeneous and inhomogeneous DNA chain) soliton propagation turns out to be favored when using a Morse potential to model the pairing interaction. This is again a welcome feature: a more realistic choice for the interaction potential enhances soliton propagation.
- In the case of the more realistic inhomogeneous chains, and in particular the chain corresponding to the real DNA base sequence of the Human Adenovirus 2, the best results concerning soliton propagation are obtained using the CY model with Morse pairing potential. Twist solitons propagate for considerable distances along the chain before stopping due to phonon emission. The mean distance is 6 times the soliton diameter, which is about 60 bp, and the maximum distance reached is about 10 times the soliton diameter. This has to be compared with soliton propagation when the same DNA chain is described by the Y model with a Morse pairing potential. In this case the mean and the maximum distance reached by solitons is, respectively, 4 times and 8 times the soliton diameter. Thus, soliton propagation in real inhomogeneous DNA chains seems to favor the CY model with Morse pairing potential.
- Soliton dynamics appears to be substantially equivalent for a real DNA base sequence and for a random one. This means that the DNA base sequence, which is of course fundamental for biological processes, does not play a significant role (at least at the level of the mesoscopic models considered here) in the torsional dynamics of DNA. This fact is a confirmation of the results of Cuenda and Sánchez [69].
- Strong limitation of our investigation are that we have mainly considered kinks with speeds of order 1 km/s and the impossibility of performing simulations at speeds of the order of magnitude of the replication speed in real DNA (about 1000 bp/s). However, by considering the motion of kinks in DNA chains subject to external forces and dissipation, we have shown that we can slow down the soliton speed to reach speeds which are only 5 order of magnitude higher than the replication speed in real DNA. This gives a strong indication that solitons can propagate in the DNA chain also with speed of order 1000 bp/s.

- Our investigation have shown that solitons propagation with average constant, non-vanishing, speed is possible also in inhomogeneous chains with real base sequences and with dissipation if an external force is present. Soliton propagation discriminates between the Morse and the harmonic potentials also when dissipative effects and external forces are present. Again, a Morse pairing potential favors soliton propagation.
- In our investigation we have not considered the effect of random thermal effects. It is therefore an open question whether soliton propagation is robust upon the inclusion of these effects. Previous numerical investigations of statistical mechanics of DNA models have shown that big excitations of the DNA chain can grow by collecting the energy of small ones [52].
- As a byproduct, our numerical investigation has shown that soliton propagation is also possible in highly inhomogeneous media. This possibility can be traced back to two different features of our mechanical system. The first is the presence in the molecular chain of both a homogeneous part that supports the topological soliton (the sugar-phosphate group) and an inhomogeneous part (the bases) that plays the role of a dissipative medium. The second is that the Morse potential localizes the interaction of the inhomogeneous part essentially near the potential minimum; away from this minimum the interaction becomes very weak: again, this weakens the soliton sensitivity to inhomogeneities in the chain.

Acknowledgments

We thank very much Giuseppe Gaeta for many enlightening discussions. We are also very grateful to Ernst Hairer for his assistance on the symplectic integrators software and, in particular, for kindly providing us his implementation of the Hamiltonian version of the software.

References

[1] M. Cadoni, R. De Leo and G. Gaeta, A composite model for DNA torsion dynamics, *Phys. Rev. E* **75** (2007) 021919, q-bio/0604014.
[2] A. Davydov, *Solitons in Molecular Systems* (Kluwer, Dordrecht, 1981).
[3] S. Englander, N. Kallenbach, A. Heeger, J. Krumhansl and A. Litwin, *Proc. Nat. Acad. Sci. USA* **77** (1980) 7222–7226.
[4] L. Yakushevich, *Nonlinear Physics of DNA* (Wiley, 1998).
[5] M. Peyrard, *Nonlinearity* **17** (2004).
[6] M. Peyrard, S. Cuesta-Lopez and G. James, *J. Biol. Phys.* **35** (2009) 73–89.
[7] S. Cuesta-Lopez, J. Errami and M. Peyrard, *J. Biol. Phys.* **31** (2005) 273–301.
[8] N. Bruant, D. Flatters, R. Lavery and D. Genest, From atomic to mesoscopic descriptions of the internel dynamics of DNA, *Biophysical Journal* **77** (1999) 2366–2376.
[9] M. Peyrard, A. Bishop and T. Dauxois, Dynamics and thermodynamics of a nonlinear model for DNA denaturation, *Phys. Rev. E* **47** (1992).
[10] A. Campa, Bubble propagation in a helicoidal molecular chain, *Phys. Rev. E* **63** (2000) 021901.
[11] M. Cadoni, R. De Leo and G. Gaeta, *J. Nonlinear Math. Phys.* **14** (2007).
[12] M. Cadoni, R. De Leo and G. Gaeta, Sine-gordon solitons, auxiliary fields, and singular limit of a double pendulums chain, *J. Phys. A* **40** (2007) 12917–12929.
[13] M. Cadoni, R. De Leo, S. Demelio and G. Gaeta, Nonlinear torsional dynamics of DNA: Some recent results, *Internat. J. Non-Linear Mech.* (2008).
[14] L. Yakushevich, Nonlinear dynamics of DNA: Velocity of the kinks activated in homogeneous polynucleotide chains, *Internat. J. Quantum Chem.* **110** (2010) 270–275.
[15] W. Greenleaf, M. Woodside and S. Block, *Ann. Rev. Biophys. Mol. Struct.* **36** (2007) 171–190.

[16] D. S. J. Yan and J. Marko, A composite model for DNA torsion dynamics, *Phys. Rev. E* **70** (2004) 011905.
[17] S. Smith, L. Finzi and C. Bustamante, *Science* **258** (1992) 1122–1126.
[18] L. F. S. B. Smith and C. Bustamante, *Science* **271** (1996) 795.
[19] J. Allemand, D. Bensimon, R. Lavery and V. Croquette, *PNAS USA* **95** (1998) 14152.
[20] T. Strick, V. Croquette and D. Bensimon, *Nature* **404** (2000) 901.
[21] H. Zhou, Y. Zhang and Z. Ou-Yang, *Phys. Rev. E* **62** (2000) 1045.
[22] A. Sarkar, J. Leger, D. Chatenay and J. Marko, *Phys. Rev. E* **63** (2001) 051903.
[23] P. Thomen, U. Bockelmann and F. Heslot, *Phys. Rev. Lett.* **88** (2002) 248102.
[24] R. Lavery, A. Lebrun, J. Allemand, D. Bensimon and V. Croquette, *J. Phys. C* **14** (2002) R383–R414.
[25] T. Strick, M. Dessinges, G. Charvin, J. A. N. H. Dekker, D. Bensimon and V. Croquette, *Rep. Prog. Phys.* **66** (2003) 1–45.
[26] J. G. Z. Bryant, M. Stone, M. Nollmann, N. Cozzarelli and C. Bustamante, *Nature* **439** (2006) 100.
[27] F. Ritort, *J. Phys. C* **18** (2006) R531–R583.
[28] H. Zhang and J. Marko, *Phys. Rev. E* **77** (2008) 031916.
[29] O. Saleh, D. McIntosh, P. Pincus and N. Ribeck, *Phys. Rev. Lett.* **102** (2009) 068301.
[30] F. Mosconi, J. Allemand, D. Bensimon and V. Croquette, *Phys. Rev. Lett.* **102** (2009) 078301.
[31] V. Baldazzi, S. Cocco, E. Marinari and R. Monasson, *Phys. Rev. Lett.* **96** (2006) 128102.
[32] C. Ke, M. Humeniuk, H. S-Gracz and P. Marszalek, *Phys. Rev. Lett.* **99** (2007) 018302.
[33] S. Zdravkovic and M. Sataric, *Phys. Rev. E* **77** (2008) 031906.
[34] Y. H. O. Ohara, A. Takatsuki, H. Itoh, N. Shimamoto and K. Kinoshita, *Nature* **409** (2001) 113.
[35] K. Adelman, A. L. Porta, T. Santangelo, J. Lis, J. Roberts and M. Wang, *PNAS* **99** (2002) 13538–13543.
[36] J. Andrecka, R. Lewis, F. Bruckner, E. Lehmann, P. Cramer and J. Michaelis, *PNAS* **105** (2008) 135–140.
[37] H. Kabata, O. Kurosawa, I. Arai, M. Washizu, S. Margarson, R. Glass and N. Shimamoto, *Science* **262** (1993) 1561–1563.
[38] E. Abbondanzieri, W. Greenleaf, J. Shaevitz, R. Landick and S. Block, *Nature* **438** (2005) 460–465.
[39] P. Thomen, P. Lopez and F. Heslot, *Phys. Rev. Lett.* **94** (2005) 128102.
[40] R. De Leo and S. Demelio, Numerical analysis of solitons profiles in a composite model for DNA torsion dynamics, *Internat. J. Non-Linear Mechanics* (2008).
[41] C. B. Tabi, A. Mohamadou and T. Kofané, Soliton excitation in the DNA double helix, *Phys. Scr.* **77** (2008) 045002.
[42] M. Peyrard and A. Bishop, Statistical mechanics of a nonlinear model for DNA denaturation, *Phys. Rev. Lett.* **62** (1989) 2755–2758.
[43] T. Dauxois, *Phys. Lett. A* **159** (1991).
[44] M. Barbi, S. Cocco and M. Peyrard, Helicoidal model for DNA opening, *Phys. Lett. A* **253** (1999) 161–167.
[45] M. Barbi, S. Cocco, M. Peyrard and S. Ruffo, *J. Biol. Phys.* **24** (1999).
[46] S. Cocco and R. Monasson, *Phys. Rev. Lett.* **83** (1999).
[47] L. Yakushevich, *Phys. Lett. A* **136** (1989).
[48] L. Yakushevich, Nonlinear DNA dynamics: Hyerarchy of the models, *Physica D* **79** (1994) 77–86.
[49] L. Yakushevich, Is DNA a nonlinear dynamical system where solitary conformational waves are possible? *J. Biosciences* **26** (2001) 305–313.
[50] G. Gaeta, On a model of DNA torsion dynamics, *Phys. Lett. A* **143** (1990) 227–232.
[51] G. Gaeta, Results and limitations of the soliton theory of DNA, *Journal of Biological Physics* **24** (1999) 81.

[52] G. Gaeta, C. Reiss, M. Peyrard and T. Dauxois, *Rivista del Nuovo Cimento* **17**(4) (1994).
[53] R. J. Roberts, *Nobel Lecture* (1993).
[54] N. Huang and A. D. MacKerell, Atomistic view of base flipping in DNA, *Phil. Trans. A* **362** (2004) 1439–1460.
[55] J. R. Horton, G. Ratner, N. K. Banavali, N. Huang, Y. Choi, M. A. Maier, V. E. Marquez, A. D. MacKerell and X. Cheng, Caught in the act: Visualization of an intermediate in the DNA base-flipping pathway induced by hhai methyltransferase, *Nucleic Acids Research* **32** (2004) 3877.
[56] C. Calladine and H. Drew, *Understanding DNA* (Academic Press, London, 1992).
[57] W. Saenger, *Principles of Nucleic Acid Structure* (Springer Verlag, 1984).
[58] L. Yakushevich, A. Savin and L. Manevitch, 66, *Phys. Rev. E* **75** (2002) physics/0204088.
[59] G. Gaeta, Solitons in yakushevich-like models of DNA dynamics with improved intrapair potential, *J. Nonlinear Math. Physics* **14** (2007) 57–81.
[60] E. Lennholm and M. Hörnquist, Revisting salerno's sine-gordon model of DNA: Active regions and robustness, *Physica D* **177** (2003) 233–241.
[61] Pdb repository.
[62] Glacton project.
[63] H. Drew, R. Wing, T. Takano, C. Broka, S. Tanaka, K. Itakura and R. Dickerson, Structure of a B-DNA dodecamer: Conformation and dynamics, *Proc. Natl. Acad. Sci. USA* **78** (1981) 2179–2183.
[64] J. G. and N. K., Principles of discrete time mechanics: I. particle systems, *J. Physics A* **30** (1997) 3115–3144.
[65] Numerical Recipes.
[66] GNU Scientific Library.
[67] E. Hairer, C. Lubich and G. Wanner, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations* (Springer Verlag, 2002).
[68] O. Braun and Y. Kivshar, Nonlinear dynamics of the Frenkel–Kontorova model, *Physics Reports* **306** (1998).
[69] S. Cuenda and A. Sanchez, Nonlinear excitations in DNA: Aperiodic models versus actual genome sequences, *Phys. Rev. E* **70** (2004) 051903.