Topological sectors of Hamiltonian cycles on periodic lattices

Paolo Glorioso*
Center for Theoretical Physics, Massachusetts Institute of Technology Cambridge, MA 02139, U.S.A.
(Dated: May 16, 2013)

We review the field theory counting of Hamiltonian cycles on periodic lattices, and provide a refinement that keeps into account their topological structure. We use this result to propose an upper bound to Hamiltonian cycles on lattices with open boundary conditions.

I. INTRODUCTION

When a single polymer chain with excluded volume interactions is brought into a bad solvent, the solvent molecules strongly repel those of the chain. To minimize the interface between solvent and chain molecules, the chain will collapse to an extremely compact configuration [1]. The study of such collapsed configurations is very relevant also because most biopolymers are compact in their native conditions. DNA, for instance, is frequently so long that it would not fit inside the cell if allowed to fluctuate freely.
The natural approach for the study of such compact configurations is to approximate the space to a lattice, and to regard compact polymers as Hamiltonian walks, i.e. paths that visit all the sites of the lattice exactly once. A pictorial representation of this approximation is shown in fig.1. The first question that one might ask is then how many Hamiltonian walks are in a given lattice, which already yields some information about the thermodynamics of the system. For large number of lattice sites, N, the number of Hamiltonian walks $\mathcal{N}_H$ is specified by the connectivity constant ω_H, which is defined by

$$\log \omega_H = \lim_{N\to\infty} \frac{\log \mathcal{N}_H}{N}. \tag{1}$$

The first estimate of ω_H goes back to Flory, where the purpose was to evaluate the configurational entropy of polymer chains immersed in a solvent [2], and it was found that

$$\omega_H = \frac{z-1}{e},$$

where z is the coordination number, i.e. the number of neighbors to a given lattice site, or equivalently the number of links attached to it. After Flory, many other estimates have been done in more refined contexts [3]. A more general approach, applicable in principle to any dimenson, was developed in 1985 by Orland et al. [4], who prescribed a systematic expansion of ω_H in terms of the coordination number for closed Hamiltonian walks, which in the literature are called Hamiltonian cycles, or Hamiltonian circuits.

*Electronic address: paolo_g@mit.edu

In the second Section of this paper we review the calculation in [4]. In the third Section we give a generalization of such calculation, which we call "weighted links", and then we use it to introduce a chemical potential associated to a discrete number that characterizes the topology of the Hamiltonian cycles. At the end of the third Section we obtain a bound on the counting of Hamiltonian cycles with open boundary condition. We conclude by discussing the validity of our calculation, and mention the way to improve it.

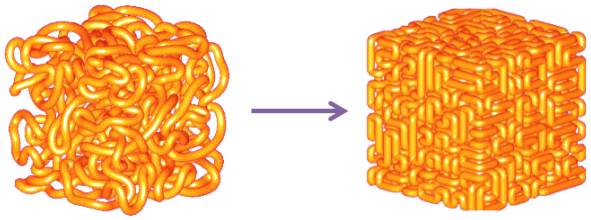

FIG. 1: Approximation of a realistic compact polymer with a Hamiltonian walk.

II. THE FIELD THEORY CALCULATION

A. Statement of the problem

Consider a lattice with coordination number z, with N sites and with periodic boundary conditions. We want to count all the Hamiltonian cycles of the lattice. The total number is equal to the following "partition function"

$$Z = \sum_{\{\mathrm{HC}\}} 1. \tag{2}$$

In [4], it is shown that the above partition function is equal to the following field theory correlator

$$Z = \lim_{n\to 0} \frac{1}{n} \frac{\int [d\varphi] e^{-\frac{1}{2}\sum_{rr'} \varphi_r \Delta^{-1}_{rr'} \varphi_{r'}} \prod_r \frac{\varphi_r^2}{2}}{\int [d\varphi] e^{-\frac{1}{2}\sum_{rr'} \varphi_r \Delta^{-1}_{rr'} \varphi_{r'}}}, \tag{3}$$

where r labels the lattice sites, φ_r is a vector of n real entries for each r, $[d\varphi] = \prod_r d\varphi_r$, and $\Delta_{rr'}$ is the connectivity matrix, i.e. $\Delta_{rr'} = 1$ when r, r' are neighbors, and $\Delta_{rr'} = 0$ otherwise. For example, for a 2-dimensional square lattice we can take $r = (m, n)$ where

$m, n = 1, \ldots, \sqrt{N}$ are integers and play the role of the coordinates of the lattice, and

$$\Delta_{rr'} \equiv \Delta_{\substack{m\,n\\ m'\,n'}} = \delta_{|m-m'|,1}\delta_{n,n'} + \delta_{m,m'}\delta_{|n-n'|,1}.$$

To prove the identity between (2) and (3), we use Wick's theorem [5] to obtain

$$Z = \lim_{n\to 0} \frac{1}{n}\frac{1}{2^N} \sum_{\text{permutations}} \langle \varphi^{i_1}_{r_1}\varphi^{i_2}_{r_2}\rangle \cdots \langle \varphi^{i_{N-1}}_{r_{N-1}}\varphi^{i_N}_{r_N}\rangle, \quad (4)$$

where the factor of $\frac{1}{2^N}$ keeps into account that there are permutations that switch $\varphi_{r_{2k-1}}$ with $\varphi_{r_{2k}}$, which should not be counted as generating new pairings since from (3) we see that $r_{2k-1} = r_{2k}$. Since the propagator of this theory $\Delta_{rr'}$ is precisely the connectivity matrix, it is easy to see that each term of the sum in (4) can be interpreted as a cycle that visits all the sites of the lattice exactly once. These cycles are precisely the Feynman diagrams of the amplitude (4). Note however that among these terms there are also disconnected cycles, which would occur when we have a combination like

$$\langle \varphi^{i_1}_{r_1}\varphi^{i_2}_{r_2}\rangle \langle \varphi^{i_{i_2}}_{r_{i_2}}\varphi^{i_1}_{r_1}\rangle \cdots,$$

so that (4) would actually be an overcounting of the Hamiltonian cycles, which are only the connected ones. The $n \to 0$ limit overcomes this issue. The propagator of this theory reads

$$\langle \varphi^i_r \varphi^j_{r'}\rangle = \delta_{ij}\Delta_{rr'}, \quad (5)$$

where $i, j = 1, \ldots, n$ are the indices of the n-dimensional "replica" vector. At a given n, we can view the corresponding partition function Z_n from (3) as summing over all possible cycles (including disconnected ones) connecting sites among n copies of the lattice. However, from (5) we see that the lines conserve the index i, which implies that there are no diagrams connecting different copies of the lattice. Therefore, connected cycles carry an overall factor of n, because there is one cycle per lattice copy, whereas cycles composed by two connected components carry a factor of n^2, and so forth. To single out the counting of the connected components only, we multiply Z_n by $\frac{1}{n}$ and take the limit $n \to 0$. Clearly, this procedure makes sense only if Z_n is analytic in n, and as shown in [4], this is indeed the case. We conclude that (2) and (3) are equal.

B. Saddle-point evaluation

It is remarkable that (2) has a field theoretic counterpart. However, it is not an easy task to evaluate (3) exactly. Following [4], we will perform a saddle-point evaluation. Note that the denominator of (3) is equal to $\exp(-\frac{n}{2}\log\det\Delta_{rr'})$, which tends to 1 as $n \to 0$. We will therefore ignore this factor. Let us write the numerator as

$$\int [d\varphi] e^{-S}, \quad (6)$$

where

$$S = \frac{1}{2}\sum_{rr'} \varphi_r \Delta^{-1}_{rr'}\varphi_{r'} - \log\prod_r \frac{\phi_r^2}{2}.$$

We shall now find the configuration ϕ_r that extremizes S. Since we imposed periodic boundary conditions, the partition function Z has the same dependence on all the site fields φ_r, and thus it is plausible to assume that the saddle-point is homogeneous, i.e. $\varphi_r = \varphi$. Therefore

$$S = \frac{1}{2}\varphi^2 \sum_{rr'} \Delta^{-1}_{rr'} - N\log\frac{\varphi^2}{2}. \quad (7)$$

The saddle-point equation is obtained by taking the first derivative of (7), i.e.

$$\sum_{rr'} \Delta^{-1}_{rr'}\varphi - N\frac{2}{\varphi} = 0. \quad (8)$$

It is easy to show that

$$\sum_{rr'} \Delta^{-1}_{rr'} = \frac{N}{z}, \quad (9)$$

indeed, from $\sum_{r'}\Delta_{rr'} = z$, we have

$$1 = \sum_s \sum_t \Delta^{-1}_{rs}\Delta_{st} = \sum_s \Delta^{-1}_{rs} z,$$

which shows (9), and therefore the solution to (8) is

$$\varphi^2 = 2z,$$

which gives

$$e^{-S} = \left(\frac{z}{e}\right)^N.$$

Note that the saddle-point equation is invariant under the $O(n)$ rotations acting on the i-index. This means that we actually have a family of saddle-points given by the action of this group, and the value of (6) is given by

$$\int [d\varphi] e^{-S} = S^n \left(\frac{z}{e}\right)^N, \quad (10)$$

where S^n is the volume of the n-sphere. Now,

$$\lim_{n\to 0}\frac{1}{n}S^n = \lim_{n\to 0}\frac{1}{n}\frac{2\pi^{n/2}}{\Gamma(n/2)} = 1,$$

and therefore we conclude

$$Z = \left(\frac{z}{e}\right)^N,$$

which gives $\omega_H = \frac{z}{e}$.
A comparison of this value with numerical results from [6] is given in fig.2. Note that the value of ω_H is better approximated by large values of N, which motivates the definition (1) of ω_H. According to [4], the value of ω_H is in very good agreement with both numerical and exact results in the literature, which seems to be due to the fact that fluctuations around the saddle-point are small.

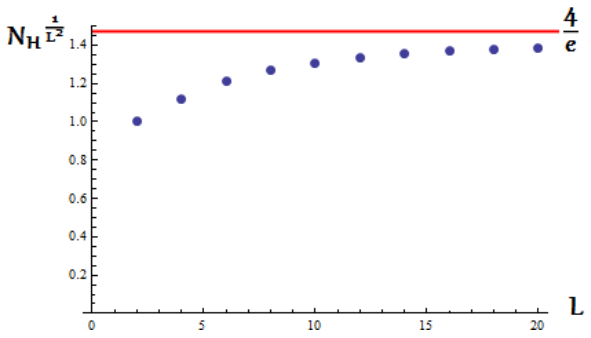

FIG. 2: Plot of $\mathcal{N}_H^{1/L^2}$ for a periodic $L \times L$ square lattice as a function of L. The red line is the value of the connectivity constant of the square obtained in this Section. As explained, we expect the dots to approximate the red line for large values of L.

III. WEIGHTED LINKS

We shall now perform a generalization of the previous result. Consider a square lattice, where links consist of the sides of the squares. Let us now promote the diagonals to be links as well. Since the diagonals are longer than the sides, in a particular problem we might need to give to diagonals a connectivity less than 1, i.e. they occur less often than sides, and we can imagine them to be "weighted" by a factor $z < 1$. This means that every time a diagonal is used to compose the cycle, the term corresponding to such cycle in the counting (4) is multiplied by an additional power of z. The situation is illustrated in fig.3. In general, we can consider a lattice with more weighted links whose connectivities are z_i, with $i = 1, \ldots, m$, where m is the number of sites connected to a given site. Repeating the field theory calculation performed above, and using the correspondingly modified connectivity matrix $\tilde{\Delta}_{rr'}$, we easily see that

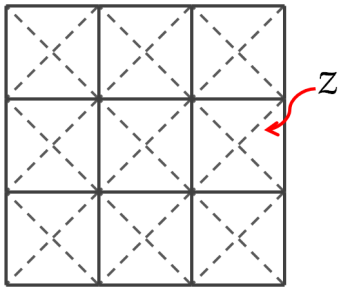

FIG. 3: The diagonals are now regarded as links with weight z.

$$\sum_{r'} \tilde{\Delta}^{-1}_{rr'} = \frac{1}{z_1 + \cdots + z_m}.$$

From this it is immediate to find the value of the partition function in the saddle-point approximation, which gives

$$\omega_H = \frac{z_1 + \cdots + z_m}{e}.$$

A. Topological sectors

Let us consider for simplicity a square lattice of size L, i.e. with L sites per side, and again with periodic boundary conditions. This lattice can be seen as a torus, and cycles wrapping around it can be classified with respect to two winding numbers $p, q \in \mathbb{Z}$ that indicate how many times a cycle winds around each of the two circles whose Cartesian product gives the torus, as depicted in fig.4.

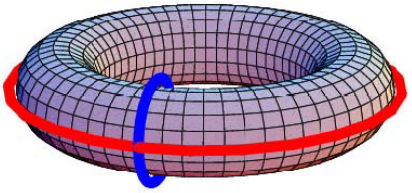

FIG. 4: Two cycles wrapping around the torus. Up to convention choice, we can say that the blue cycle has winding numbers $(p, q) = (1, 0)$, and the red cycle has winding numbers $(p, q) = (0, 1)$.

Consider the following modification of eq. (2),

$$Z = \sum_{\{\mathrm{HC}\}} e^{\mu p + \nu q}\, 1 \equiv \sum_{p,q} e^{\mu p + \nu q} Z_{p,q}, \qquad (11)$$

where $Z_{p,q}$ is the number of Hamiltonian cycles with winding numbers (p, q). It turns out that using weighted links we can estimate $Z_{p,q}$ and therefore refine the counting of the Hamiltonian cycles given by [4]. Let us introduce directions x and y in the lattice. We make the convention choice that the winding numbers p, q are referred to the winding along directions x and y, respectively, as shown in fig.5. The claim is that (11) is equivalent to

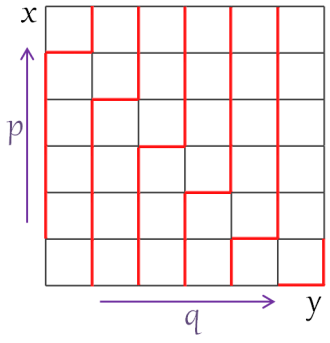

FIG. 5: Winding directions associated to p and q. The Hamiltonian cycle has $(p, q) = (5, 1)$, and $N_c = 12$.

$$\tilde{Z} = \lim_{n \to 0} \frac{1}{n} \frac{\int [d\varphi] e^{-\sum_{rr'} \varphi^*_r \tilde{\Delta}^{-1}_{rr'} \varphi_{r'}} \prod_r \varphi^*_r \varphi_r}{\int [d\varphi] e^{-\sum_{rr'} \varphi^*_r \tilde{\Delta}^{-1}_{rr'} \varphi_{r'}}}, \qquad (12)$$

where now we consider φ_r to be complex, $[d\varphi] = \prod_r d\varphi^*_r d\varphi_r$, and, writing the site label in terms of the

integer coordinates on the square $r = (m, n)$, we can define

$$\tilde{\Delta}_{rr'} \equiv \tilde{\Delta}_{\substack{m\,n\\ m'n'}} = \\ = (\delta_{|m-m'|,1}\delta_{n,n'} + \delta_{m,m'}\delta_{|n-n'|,1}) e^{i\frac{\mu}{L}(m'-m)+i\frac{\nu}{L}(n'-n)}, \tag{13}$$

i.e. $\tilde{\Delta}_{rr'}$ is equal to the connectivity matrix $\Delta_{rr'}$ with an additional factor that can be interpreted as a weight, which will eventually keep into account the winding numbers. Note that the weights are complex, but the propagator is Hermitian. The contraction of fields now becomes

$$\langle \phi_r^{i*} \phi_{r'}^{j} \rangle = \langle \phi_{r'}^{j} \phi_r^{i*} \rangle = \delta_{ij} \Delta_{rr'} e^{i\frac{\mu}{L}(m'-m)+i\frac{\nu}{L}(n'-n)}, \tag{14}$$

and using Wick's theorem, it is easy to see that the effect of this correction is to add an overall factor of $e^{i\mu p + i\nu q}$ to a given permutation of the contractions. Since each of these permutations corresponds to a Feynman diagram, which in turn corresponds to a Hamiltonian cycle, such overall factor will be the weight (up to an i in the exponent) that multiplies the "1" in (11). We then conclude that (11) is equal to (12), up to an extra i that multiplies the chemical potentials. Note that the technical reason why we introduced complex fields φ_r is to allow a non-symmetric propagator (14), which is necessary in the above construction. On the other hand, according to the Feynman rules of this theory, the complexification of φ_r introduces an orientation in the graph of the propagator, which in turn can be seen as introducing an orientation in the Hamiltonian cycles. In order to keep track of the winding numbers we do need an orientation, being these numbers of either sign, and therefore the complexification of fields is not just a trick for avoiding a technical issue: it's a conceptually necessary feature of the theory. Performing the saddle-point method on (12) gives, using the notation from Section II.B,

$$\tilde{Z} = \int [d\varphi] e^{-S} = 2S^n \left(\frac{e^{\frac{i}{L}\mu} + e^{-\frac{i}{L}\mu} + e^{\frac{i}{L}\nu} + e^{-\frac{i}{L}\nu}}{e} \right)^N,$$

where comparing to (10) we see an additional factor of 2, which comes from the complex integration.[1] Note that we expect a doubling of the counting as cycles have now an orientation, and so each of them is counted twice. The correlator is

$$\tilde{Z} = 2 \left(\frac{e^{\frac{i}{L}\mu} + e^{-\frac{i}{L}\mu} + e^{\frac{i}{L}\nu} + e^{-\frac{i}{L}\nu}}{e} \right)^{L^2}.$$

and therefore

$$Z = 2 \left(\frac{e^{\frac{1}{L}\mu} + e^{-\frac{1}{L}\mu} + e^{\frac{1}{L}\nu} + e^{-\frac{1}{L}\nu}}{e} \right)^{L^2},$$

[1] More precisely, it's the Jacobian that appears in $dzdz^* = 2dxdy$, where $z = x + iy$.

Note that setting $\mu = \nu = 0$ we recover the result from Section II, which indeed can be also computed by directly considering complex fields φ_r. We can now expand Z in order to recover the structure of (11). For simplicity, we will turn off one of the two chemical potentials, i.e. we shall consider

$$Z = 2 \left(\frac{e^{\frac{1}{L}\mu} + e^{-\frac{1}{L}\mu} + 2}{e} \right)^{L^2}.$$

Expanding,

$$Z = 2 \left(\frac{2}{e} \right)^{L^2} \sum_{m=-L^2}^{L^2} \sum_{j=0}^{L^2} \frac{L^2! 2^{-(m+2j)}}{j!(m+j)!(L^2-(m+2j))!} e^{\mu \frac{m}{L}}. \tag{15}$$

Note that the powers of e^μ run from $-L$ to L, which is what we expect as the cycles have length L^2, so that they cannot reach a winding number greater than L. However, contrary to its original definition in (11), the powers of e^μ in (15) are fractional as they are multiples of $1/L$. Since (12) is exact, we expect this fact to be an artifact of the saddle-point approximation, as will be argued in Section III.C. From (15) we have

$$Z_p = 2 \left(\frac{2}{e} \right)^{L^2} \sum_{j=0}^{L^2} \frac{L^2! 2^{-(pL+2j)}}{j!(pL+j)!(L^2-(pL+2j))!} = \\ \approx \frac{2}{\sqrt{\pi}} \left(\frac{4}{e} \right)^{L^2} e^{-p^2}, \tag{16}$$

where we approximated Z_p with a Gaussian distribution as we are considering large values of L. The standard deviation of this distribution is $\frac{1}{\sqrt{2}}$, meaning that most of the cycles have winding number between -1 and +1. This is completely reasonable, for the following consideration. Recall the computation of the semiflexible polymer done in [7],

$$Z = \sum_{\{\mathrm{HC}\}} e^{-\mu N_c} = \left(\frac{2(1+e^{-\mu})}{e} \right)^{L^2}, \tag{17}$$

where in [7] $\mu = \beta\varepsilon_h$ is a chemical potential associated to the number of corners N_c, i.e. the number of sites at which contiguous links are not parallel. Using (17) we can compute the average number of corners and its variance,

$$\langle N_c \rangle = \frac{L^2}{2}, \qquad \langle N_c^2 \rangle_c = \frac{L^2}{4}. \tag{18}$$

Now, having a winding number which is order L requires the cycle to be stretched along the direction corresponding to such large winding number, and thus N_c should be $O(L)$, as is illustrated by the example in fig.5. From (18) we know that almost all the Hamiltonian cycles have an

order L^2 number of corners, and thus we do not expect many cycles to have high winding numbers. In particular, because the average number of corners is comparable to the length of the cycle, we expect most of the cycles to lie approximately within one fundamental square (i.e. to cover the torus only once). Therefore we find it reasonable that the standard deviation given in (16) is small and independent of L.
Note that our calculation relies on the same approximations as in [4]. Therefore, for large L and large coordination number (for example in the case of a high dimensional torus), we expect this estimate to be a very good approximation.

B. Open boundary conditions

There is an interesting by-product of the result we found above, which might be more relevant for realistic biological systems. We can get an upper bound for the estimate of Hamiltonian cycles on the square with open (i.e. non periodic) boundary conditions that holds within the regime of the saddle-point approximation. Since the Hamiltonian cycles on the open square have zero winding number, we expect them to be less than the Hamiltonian cycles with zero winding number on the torus. Note that there is a technical complication, as in our approximation of Z_p, to be sure that we essentially count all the cycles with zero winding number we need to consider also fractional values of p, because Z_0 might not be the total number of trivial cycles. From (16), we see that a reasonable approximation for this number is Z_0 times twice the standard deviation and divided by 2 in order to discard orientation, i.e. $\sqrt{2}Z_0$, which then gives

$$\text{HC's on the open square } < \sqrt{\frac{2}{\pi}} \left(\frac{4}{e}\right)^{L^2} .$$

We expect that developing further the setup we gave above we can considerably lower the bound, as there are a few refinements one can do. In particular, note that this bound is given by considering the Hamiltonian cycles with $p = 0$ but with any value of q. Setting also $q = 0$ will clearly lower the bound. However, due to the fractionality of p in our evaluation, it is not clear how to safely estimate the number of Hamiltonian cycles with $(p, q) = (0, 0)$, although we expect that the evaluation of the fluctuations will give us the answer to this question. In fig.6 we compare the plot in fig.2 with the bound found above.

C. Blurred topology

It is worth to state neatly what are the limitations of the above calculation. Eq. (12) is, up to some factors of i, the exact "gran partition function" introduced in (11). This means that if we expand it in powers of the two fugacities, we obtain an integer power expansion. This property can also be seen directly by applying Wick's theorem to (12). However, in (15) we find that the powers are multiples of $1/L$, and thus are fractional. The explanation of this should come from the fact that we evaluated (12) through a saddle-point approximation. The saddle point is insensitive to the structure of the topological sectors, and therefore it mixes the contribution of cycles coming from different sectors, resulting in a "blurred topology". Said differently, let's call $\mathcal{N}_{p,q}$ the exact number of Hamiltonian cycles with winding numbers (p, q). Instead of $\mathcal{N}_{p,q}$ we obtain $Z_{p,q}$, which receives contributions from all the topological sectors. However, we expect the contributions from the sector (p', q') to $Z_{p,q}$ to be smaller as (p', q') gets further from (p, q), with respect to some notion of distance. Perhaps, by performing a computation of the fluctuations, we can obtain the behavior of these contributions in terms of (p, q, p', q'), or at least we could quantify the uncertainty on $\mathcal{N}_{p,q}$. Such computation should not be essentially different from the one performed on [4], and thus we expect it to be feasible. As we mentioned earlier, knowing the uncertainty on $Z_{0,0}$ is important as it would allow us to considerably improve the bound given above, and to quantify the uncertainty on such bound.

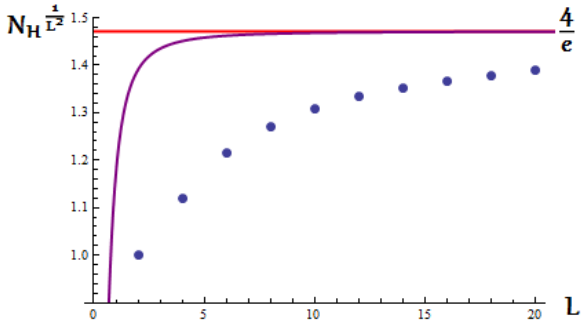

FIG. 6: The plot is a rescaled version of fig.2. The purple line corresponds to the bound.

IV. CONCLUSIONS

The field-theoretic counting given by Orland et al. in [4] has been reviewed, and by introducing the concept of weighted links we obtained a refinement of the counting by partitioning the Hamiltonian cycles into topological sectors. This refinement is applicable at least to d-dimensional cubic lattices, although we expect a much wider applicability. From the estimate of the number of trivial cycles we then inferred an upper bound to the number of Hamiltonian cycles for the square lattice with open boundary conditions, which is more similar to realistic biological systems. These evaluations hold better and better as we increase the number of lattice sites and the coordination number. The next natural step is to

evaluate the quadratic fluctuations of the winding number partition functions, which would provide a precise estimation of the uncertainty in such quantities, and moreover it would yield a considerably more accurate bound for the open boundary conditions counting. A further, more ambitious task, is to try to extend the above procedure to three-dimensional knots, which are of more basic importance in biology. The aim in this case would be to introduce a chemical potential associated to the value of the Alexander polynomial. At the moment we have no hint on whether this is possible, but there might be a chance that one can introduce a chemical potential related to quantities like the sum of the self-crossings of the projected three-dimensional walks, which would already be very interesting.

I am grateful to Professor Mehran Kardar and to Dr. Alkan Kabakçıoğlu for the various helpful discussions.

[1] P.-G. de Gennes, Scaling Concepts in Polymer Physics, Cornell University, New York (1979).

[2] P.J. Flory, Principles of Polymer Chemistry, Cornell University, New York (1995).

[3] A. Jaeckel, J. Sturm and J. Dayantis, Statistics of confined chains. III: Hamiltonian paths, 1997 J. Phys. A: Math. Gen. 30 2345.

[4] H. Orland, C. Itzykson and C. de Dominicis, An evaluation of the number of Hamiltonian paths, J. Physique Left. 46 (1985) L-353 - L-357.

[5] M. Srednicki, Quantum Field Theory, Cambridge University, New York (2007).

[6] J.L. Jacobsen, Exact enumeration of Hamiltonian circuits, walks, and chains in two and three dimensions, 2007 J. Phys. A: Math. Theor. 40 14667.

[7] S. Doniach, T. Garel, and H. Orland, Phase diagram of a semiflexible polymer chain in a θ solvent: Application to protein folding, J. Chem. Phys. 105, 1601 (1996).