On 4-point correlation functions in simple polymer models — **Source link** ↗

Johannes-Geert Hagmann, Karol K. Kozlowski, Nikos Theodorakopoulos, Michel Peyrard

Related papers:

- General formulation of long-range degree correlations in complex networks.

- Rate equation approach for correlations in growing network models.

- Characterizing the intrinsic correlations of scale-free networks

- Clustering and Dispersion Rates for Some Interacting Particle Systems on $\mathbb{Z}$

- Renormalization group analysis of the small-world network model

On 4-point correlation functions in simple polymer models

Johannes-Geert Hagmann, Karol Kozlowski, Nikos Theodorakopoulos, Michel Peyrard

On 4-point correlation functions in simple polymer models

Johannes-Geert Hagmann[1,‡], Karol K. Kozlowski[1,‡], Nikos Theodorakopoulos[2] and Michel Peyrard[1]

[1] *Université de Lyon, Laboratoire de Physique, Ecole Normale Supérieure de Lyon,*
CNRS, 46 Allée d'Italie, 69364 Lyon, France and

[2] *Theoretical and Physical Chemistry Institute, National Hellenic Research Foundation,*
Vasileos Constantinou 48, 116 35 Athens, Greece

(Dated: April 17, 2009)

We derive an exact formula for the covariance of cartesian distances in two simple polymer models, the freely-jointed chain and a discrete flexible model with nearest-neighbor interaction. We show that even in the interaction-free case correlations exist as long as the two distances at least partially share the same segments. For the interacting case, we demonstrate that the naive expectation of increasing correlations with increasing interaction strength only holds in a finite range of values. Some suggestions for future single-molecule experiments are made.

I. INTRODUCTION

With the advent of new powerful tools for imaging and manipulation, the detection of single molecules and the characterization of their dynamics is a rapidly growing and developing field in experimental biophysics. Recently, the possibility to track molecular motion at the level of a single molecule has received substantial attention. Indeed, such methods can in principle provide a glimpse of a biological molecule "at work" (see [1] for a recent review). The complexity and variety of interactions in biomolecules makes it hard to study quantitatively the biomolecular dynamics through a theoretical model that ought to reproduce most of the experimental observations. It is thus in computer simulations of empirical models [2, 3] that the most considerable quantitative studies of biomolecular dynamics can be found. The challenge of analyzing simulation results is usually somewhat opposite to the case of single-molecule experiments. On the one hand, in experiments one attempts to extract as much information on the dynamics as possible from a small number of accessible observables. On the other hand, for the interpretation of numerical simulations one needs to develop schemes allowing to reduce a large amount of data resulting from the numerous degrees of freedom to a meaningful subset.

Obviously, if not guided by prior motivation, the ways to achieve this reduction are manifold. A large number of such schemes have been proposed, involving linear and nonlinear concatenation of degrees of freedom (see e.g. [4] for a comparison of common approaches). Unfortunately, very few of these methods provide testable predictions on observables that are directly accessible in experiments. Hence, their potential for the interpretation of experiments probing biomolecular motion is limited.

In this work, we explore an alternative direction for probing biomolecular motion by introducing an experimentally testable and hence confutable framework allowing to characterize cooperative motion in the simulation of empirical models for biomolecules. More precisely, we introduce 4-point correlation functions of distances as a new measure that can serve to compare results from multiple dye single-molecule Förster-Resonance Energy Transfer (FRET) with simulations and analytical results on models of biopolymers. While most measures characterizing the dynamics of biomolecules start directly from numerical observations, in this work we build the theory of 4-point correlation functions in the limiting cases of exactly solvable polymer models which are subsequently validated by numerical experiments. As the implementation of the formalism is not limited to the underlying model, applications to more complex models which can only be treated numerically will be reported in future work. The results on the simple models presented here already reveal important issues regarding the influence of interaction strength as well as the geometry that are relevant to the interpretation of future experiments.

The article is organized as follows. In section II we outline the interest of choosing a 4-point correlation function for characterizing the collective motion of biomolecules as a generalization of the present 2-point FRET-based experiments. Focussing on analytically solvable models, we calculate both analytically and numerically the 4-point correlations functions of two simple polymer models: the freely-jointed chain in section III and a discrete flexible model with nearest-neighbor interaction in section IV. In both cases we study the convergence of covariances as estimated in computational studies or single-molecule experiments from finite-time averages towards the ensemble average determined by our exact calculations. We conclude by a discussion of our findings in respect to future single molecule experiments in section V.

---

‡The two authors have equally contributed to this work.

II. DESIGN OF 4-POINT CORRELATION FUNCTIONS FOR BIOMOLECULES

Present day single-molecule FRET techniques can retrieve the time-resolved distance between a donor and acceptor molecule. The donor/acceptor is a dye molecule attached to the biomolecule or an unit of the biomolecule itself, whose dynamics reveal the motion of the underlying molecule. Recent highlights in experimental single molecule FRET setups include the high resolution long-time scale observation of protein/ligand and protein dynamics [5, 6], and the observation of fluctuations by three-color FRET in the DNA four-way junction [7]. The measurement of FRET signals with more than two dye molecules in a single biomolecule has the advantage of yielding more than one distance at a time. Still, such approaches are technically difficult to realize, and require to overcome several difficulties, e.g. the selective dye labelling within the probe molecule and a more demanding theoretical framework to extract distance information from the multiple efficiency signals [8] to name but a few.

Despite these experimental and theoretical challenges, it is constructive to anticipate what information is to be gained from such experiments, keeping in mind the rapid progress in this field over the past years. We address the question: how do *two* distances revealed by FRET measurement correlate in biomolecules? In other words, given two distinct pairs of points (A, B) and (C, D) in the single molecule, to what extent can correlation be expected when one meaures the time-dependent distances $R_1 = |\mathbf{R}_{AB}|$ and $R_2 = |\mathbf{R}_{CD}|$?

The covariance of distances constitutes a simple measurement of this correlated motion. For the two time-dependent observables, its definition reads

$$\mathcal{C}(R_1^2, R_2^2, T) = \overline{\left(R_1^2(t) - \overline{R_1^2} \right)\left(R_2^2(t) - \overline{R_2^2} \right)} = \overline{R_1^2(t) R_2^2(t)} - \overline{R_1^2}\,\overline{R_2^2} \quad . \tag{1}$$

There $\overline{A(t)} = \frac{1}{T}\int_{t_0}^{t_0+T} dt\ A(t)$ denotes the time average of the observable $A(t)$ over a time period T. The average is assumed to start at the beginning of the measurement when $t = t_0$, consequently, the value of $\mathcal{C}(R_1^2, R_2^2, T)$ in principle also depends on the choice of t_0. This dependence only vanishes in the limit of long T or averaging over different t_0. $\mathcal{C}(R_1^2, R_2^2, T)$ is positive if the two signals simultaneously increase/decrease and negative if they vary in the opposite sense. Finally, the covariance is zero if the fluctuations of these distances are uncorrelated. We further denote by $C(R_1^2, R_2^2)$ the ensemble average value of the covariance to which $\mathcal{C}(R_1^2, R_2^2, T)$ should converge in the limit $T \to \infty$. We will moreover consider a reduction of the experimental setting, where the maximum distance for which a FRET signal can be obtained is limited, and calculate the covariance for any distance.

In the following sections, we calculate the covariance $C(R_1^2, R_2^2)$ analytically and numerically for two simple polymer models, the freely-jointed chain which is the simplest model, and a discrete flexible model which corresponds to polymer chains with nearest-neighbor interactions.

III. FREELY-JOINTED CHAIN

Let us now derive the distance covariance $C(R_1^2, R_2^2)$ from ensemble averages for the freely jointed chain representing a classical polymer model without interactions. We consider a chain of N monomers of fixed length b with no explicit interactions between them. Each monomer vector $\mathbf{r}_i$ ($|\mathbf{r}_i| = b$) is labelled by a discrete index. We will denote the distances to be considered for correlation by $\mathbf{R}_1 = \sum_{i=k_1}^{k_2} \mathbf{r}_i$, $\mathbf{R}_2 = \sum_{i=k_3}^{k_4} \mathbf{r}_i$, where the order $k_1 \leq k_2$, $k_3 \leq k_4$, $k_i \in \{1, 2, ..., N\}$ is assumed. This notation is illustrated in figure 1. Most of the classical results on this chain can be found in [9]. The N-segment probability distribution function factorizes due to the lack of explicit interaction. It reads

$$P(\{\mathbf{r}_i\}) \ = \ \Pi_{i=1}^{N} p(\mathbf{r}_i) = \Pi_{i=1}^{N} \frac{\delta(|\mathbf{r}_i| - b)}{4\pi b^2} \quad . \tag{2}$$

A.   Analytical results

The purpose of this calculation is to understand how non-vanishing 4-point correlations can arise between distances in polymers even if there is no specific interaction that could be responsible for cooperative effects. We study a correlator between the squared distances

$$C(R_1^2, R_2^2) \ \equiv \ \langle \mathbf{R}_1^2 \mathbf{R}_2^2 \rangle - \langle \mathbf{R}_1^2 \rangle \langle \mathbf{R}_2^2 \rangle \quad . \tag{3}$$
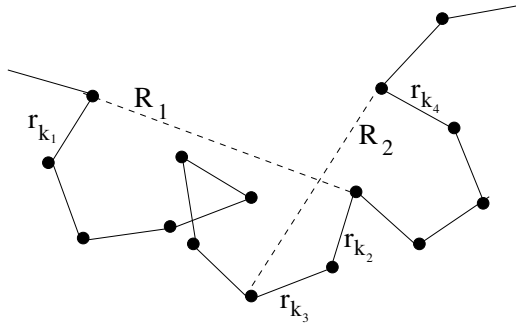
FIG. 1: scheme of the notations used for the freely jointed chain

As the probability distribution function factorizes, one has $C\left(R_1^2, R_2^2\right) = 0$ whenever $\{k_1, ..., k_2\} \cap \{k_3, ..., k_4\} = \emptyset$. Consequently, we only consider the case $\{k_1, ..., k_2\} \cap \{k_2, ..., k_4\} \neq \emptyset$. We split the summation in a way allowing to cancel the factorizable terms:

$$\langle \mathbf{R}_1^2 \mathbf{R}_2^2 \rangle = \langle \left[\sum_{ij=k_1}^{k_3-1} (\mathbf{r}_i \cdot \mathbf{r}_j) + \sum_{ij=k_3}^{k_2} (\mathbf{r}_i \cdot \mathbf{r}_j) \right] \cdot \left[\sum_{kl=k_3}^{k_2} (\mathbf{r}_k \cdot \mathbf{r}_l) + \sum_{kl=k_2+1}^{k_4} (\mathbf{r}_k \cdot \mathbf{r}_l) \right] \rangle . \tag{4}$$

There, the non-factorizable contribution arises from the overlapping sums.

Summing up the factorizable terms, we have

$$C(R_1^2, R_2^2) = \langle \left(\sum_{ij=k_3, \, i\neq j}^{k_2} (\mathbf{r}_i \cdot \mathbf{r}_j) \right)^2 \rangle = \frac{2}{3} b^4 (k_2 - k_3 + 1)(k_2 - k_3) , \tag{5}$$

after an explicit integration in spherical coordinates. From this simple result, we draw the following conclusions.
i) A non-zero covariance arises for geometrical overlap in the distances R_1, R_2 along the sequence, i.e. if R_1 and R_2 share at least <u>two</u> monomers, the covariance of the squared distances scales as the square of the overlap $\zeta = k_2 - k_3 + 1$. The result is independent of the size of the system, or the overall length of the distances R_1, R_2 themselves as it should be. The distances have to share at least two monomers as for as single shared monomer the auto-correlation in $\langle R_1^2 R_2^2 \rangle$ is still contained in the offset $\langle R_1^2 \rangle \langle R_2^2 \rangle$. From a physical point of view, the result is plausible as no information from neighboring segments can be transmitted through the chain owing to the absence of interacting forces. Only the part of the topology which is shared among the distances R_1, R_2, i.e. the ζ overlapping segments, can contribute to a non-vanishing correlation.
ii) The value of the covariance is positive semidefinite; there is on average no anti-correlated motion in the chain.
In the interacting case this picture should change. In particular, the result should involve the whole lengths R_1, R_2 at least for moderate interactions. Yet, strong interaction will tend to constraint the motion of the polymer into a single configuration. In this case, $\langle R_1^2 R_2^2 \rangle = \langle R_1^2 \rangle \langle R_2^2 \rangle$, and $C(R_1^2, R_2^2) = 0$.

B. Numerical results

a. ensemble averages We verified (5) by simple sampling on an ensemble average of N_c randomly generated freely-jointed chains according to the probability distribution (2) with a fixed number of total segments N and fixed monomer length b. The left-hand side of figure 2 shows a comparison of the covariance values $C(R_1^2, R_2^2)$ between the ensemble averages obtained from $N_c = 5 \cdot 10^4$ chains of length $N = 50$ segments and the analytical result for varying number of overlap monomers $\zeta = 0, 1, ..., 19$. The spacing between the points increases as the covariances scales with ζ^2 in the overlap.
b. time averages Besides a static ensemble average, we can also estimate dynamical time averages. This can be achieved by implementing a chain of mass points with rigid links for which we numerically solve the constrained equations of motion. An adequate representation for the freely-jointed chain can be obtained by simulating a chain of N harmonic springs using constrained molecular dynamics. Here, we use the popular Brooks-Brunger-Karplus algorithm for Langevin dynamics [10] combined with RATTLE [11] to solve the equations of motion of $N+1$ particles joined by harmonic springs, applying the constraints $|\mathbf{r}_i| = b$ at each time-step of the integration. Note that such an approach has to be taken with care for several reasons.
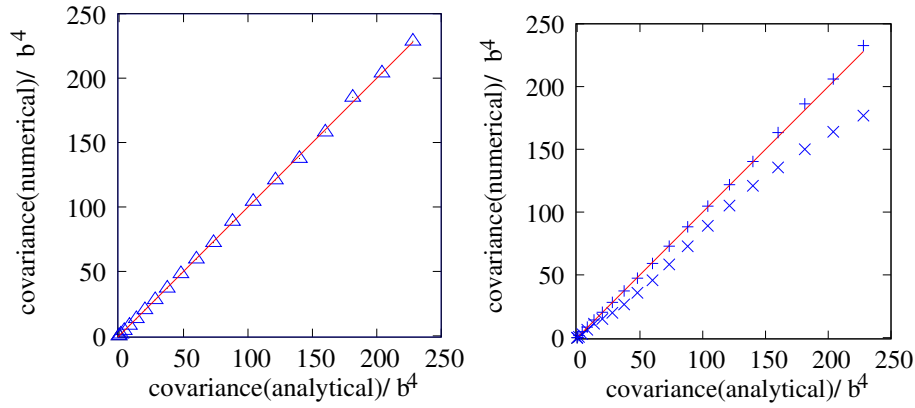
FIG. 2: *Left:* Covariance calculated according to (5) and numerical results (simple sampling, $N_c = 5 \cdot 10^4$, $N = 50$, $\zeta = 0, 1, ..19$). The straight line $f(x) = x$ is a guide to the eye. *Right:* Covariance calculated according to (5) and numerical results (molecular dynamics, see text). $+$: $m = 1.0$, $\times$: $m = 100.0$, $\zeta = 0, 1, ..19$.

i) Ideally, such a chain would be massless to avoid biasing the trajectory of an individual monomer by a collective inertial motion. This case can only be approximated if the chain is strongly coupled to a heat bath so that inertial effects are overdamped, and the time interval between the configurations recorded to calculate averages is sufficiently long so that the monomer orientations can be effectively considered as uncoupled.

ii) It has recently been pointed out that applying constraints to a stochastic heat bath leads to temperature dependent equilibrium bond lengths [12]. If the covariance was to be evaluated at different temperatures, as the bond length enters with b^4, a correction needs to be applied to the integration algorithm.

All observables of the simulation are expressed in dimensionless units. We simulate $N + 1 = 51$ point masses of equal mass $m = 1.0$ joint by harmonic springs (uniform spring constant $D = 1.0$) and an equilibrium bond length between two masses $b = 1.0$. Using a time step of $\Delta t = 0.1$ and a friction constant $\gamma = 0.1$, we have $\gamma \Delta t \ll 1$ and $\Delta t \ll (2\pi)/\sqrt{D/m}$. The initial velocities were taken from a Maxwell-Boltzmann distribution, and the chain is thermalized at $k_B T = 0.05$ for 10^5 time units in order to guarantee the loss of memory of the initial condition (a partially stretched configuration). Configurations were recorded every 100 time units to compute averages. For a constant temperature trajectory of length 10^6 time units, the temporally averaged kinetic energy closely approached the canonical expectation value. The right-hand side of figure 2 illustrates the effects of inertia on the comparison with the analytical results. For small masses, the overdamped dynamics of the constrained harmonic chain approximate the phase space sampling of the freely-jointed chain quite well at least on the timescale of the simulation. However, the results for a larger mass exhibits a strong deviation. The latter is observed to be more pronounced when the overlap ζ is located at the center of the chain (data not shown) compared to the case of an overlap located at the ends. Indeed, the free end is only affected by inertial motion from one side.

C. Relation of the distance covariance to the third order susceptibility in spin systems

In classical spin systems, fourth-order correlation functions arise naturally when analyzing higher order susceptibilities as response functions to an external field. To see the analogy with the calculation of the higher-order correlations with polymers, we define the Hamiltonian of the system to be

$$H = H_0 + \mathbf{h}_1 \cdot \mathbf{O}_1 + \mathbf{h}_2 \cdot \mathbf{O}_2 \tag{6}$$

where H_0 is an unperturbed Hamiltonian of some spin chain and $\mathbf{O}_1, \mathbf{O}_2$ are two observables of the system coupling to the external fields $\mathbf{h}_1, \mathbf{h}_2$. We now can introduce third-order susceptibility $\chi_{ijkl}^{(3,*)}$ associated with mixed derivatives of the on-site free energy f in the two fields $\mathbf{h}_1, \mathbf{h}_2$:

$$\chi_{ijkl}^{(3,*)} = -\frac{\partial^4 f}{\partial h_{1,i} \partial h_{1,j} \partial h_{2,k} \partial h_{2,l}} \Big|_{h_1=h_2=0} \quad , \quad f = -\lim_{N \to \infty} \frac{k_B T}{N} \log Z \ . \tag{7}$$

In this particular case we get, for Hamiltonians with vanishing first order moments in the observables $\mathbf{O}_1$ and $\mathbf{O}_2$,

$$\chi_{iijj}^{(3,*)} = (k_B T)^{-3} \lim_{N \to \infty} \frac{1}{N} \left(\langle O_{1,i} O_{1,i} O_{2,j} O_{2,j} \rangle - \langle O_{1,i} O_{1,i} \rangle \langle O_{2,j} O_{2,j} \rangle - 2 \langle O_{1,i} O_{2,j} \rangle^2 \right) \ . \tag{8}$$

$\chi_{iijj}^{(3,*)}$ can be related to $C(R_1^2, R_2^2)$. After a summation over the indices i, j, the first two terms within the bracket in the previous expression correspond to $C(R_1^2, R_2^2)$ provided that the identification $O \leftrightarrow R$ is made. The extra term $\langle R_{1,i} R_{2,j} \rangle$ has however no correspondence in the previous calculation. In contrast to the fourth-order correlation we have defined above, calculating the fourth-order susceptibility would require not only the information on the distances $|\mathbf{R}|$, but also on their orientation. Such an information cannot be accessed by FRET techniques. Hence, in practice, the measurement only allows to infer the covariance which is not an intensive quantity in the overlap.

We have seen that the analogy with spin systems provides a relation between the covariance measure with a generalized susceptibility in a spin system. For the moment, our calculations deal with the simplest possible case of a polymer chain without interaction. In the following section, we go beyond this first order, calculating the covariance in a polymer model with nearest-neighbor interactions. This allows us to assess the qualitative and quantitative change of the results upon introducing interactions. The analogy with spin systems will also prove very useful in this context, as the starting point for our calculation are exact results known for spin systems.

IV. DISCRETE FLEXIBLE CHAIN

A classical system for studying polymers is the worm-like chain model (WLC). In this section, we study the behavior of a discrete flexible chain with nearest-neighbor interactions [13] which in the continuum limit and for large coupling constants yields the WLC model. A related model has been analyzed in the context of loop formation in double stranded DNA [14]. In contrast to the freely-jointed chain which represents a non-interacting system of stiff rods, this has the additional physical feature of penalizing local bending by a harmonic force deriving from the Hamiltonian

$$H = -\epsilon \sum_{i=1}^{N-1} (\mathbf{r}_i \mathbf{r}_{i+1} - b^2) \tag{9}$$

where $|\mathbf{r}_i| = b$ and ϵ is a uniform coupling constant. A rescaling of the Hamiltonian $H \to H - \epsilon(N-1)b^2$, makes the system equivalent to the zero-field limit of the classical one-dimensional Heisenberg chain with uniform coupling and open boundaries. The latter has been extensively studied, see e.g. [15] for a physical perspective.

In 1994 Nakamura and Takahashi [16] reported the nonlinear susceptibility for a uniform field in this system. These results can thus serve as a starting point for the calculation of the distance covariance in an interacting polymer model. However, in order to be fully applicable, these should be generalized along the lines described in the following section.

A. Analytical results

As the calculation of the covariance for the discrete flexible chain is more involved than for the freely-jointed chain, we start by an outline of the procedure.

We are interested in the covariances

$$C(R_1^2, R_2^2) = \langle \mathbf{R}_1^2 \mathbf{R}_2^2 \rangle - \langle \mathbf{R}_1^2 \rangle \langle \mathbf{R}_2^2 \rangle \tag{10}$$

with $\mathbf{R}_1 = \sum_{i=k_1}^{k_2} \mathbf{r}_i$, $\mathbf{R}_2 = \sum_{i=k_3}^{k_4} \mathbf{r_i}$ as defined previously. Averages are taken with respect to the canonical probability distribution function bound to the rescaled discrete flexible chain Hamiltonian

$$H = -K k_B T \sum_{i=1}^{N-1} \hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_{i+1} \tag{11}$$

with $K = \epsilon/(k_B T b^2)$ and $\mathbf{r} = b\hat{\mathbf{r}}$. The squared distance correlation functions are deduced by using the isotropy of the zero-field Hamiltonian and Fisher's result [17] of local correlations in the model (see also Appendix A). One obtains

$$\langle \mathbf{R}_1^2 \rangle = 3\langle R_1^z R_1^z \rangle = \sum_{ij=k_1}^{k_2} u^{-|i-j|} = \frac{2(a+1) - (a+3)u^a}{1-u} - 1 - \frac{u(1-u^{a-1})}{(1-u)^2}, \tag{12}$$

where $u = \coth(K) - K^{-1}$ and $a = k_2 + 1 - k_1$. However, the correlator of the two squared distances

$$\langle \mathbf{R}_1^2 \mathbf{R}_2^2 \rangle = \sum_{ij=k_1}^{k_2} \sum_{kl=k_3}^{k_4} \langle (\mathbf{r}_i \cdot \mathbf{r}_j)(\mathbf{r}_k \cdot \mathbf{r}_l) \rangle \tag{13}$$

cannot be directly inferred from previous results and deserves special attention. The two sums run over different domains, such that the use of permutation symmetry is subject to restrictions. The isotropy of the interactions can be used to reduce the number of correlators to be evaluated to two, i.e. $\langle r_i^x r_j^x r_k^z r_l^z \rangle$ and $\langle r_i^z r_j^z r_k^z r_l^z \rangle$. All the other correlators are deduced from a rotation of the reference frame (the latter being a symmetry of the Hamiltonian). $\langle r_i^z r_j^z r_k^z r_l^z \rangle$ has been derived in [16] (see also [18] cited therein), whereas $\langle r_i^x r_j^x r_k^z r_l^z \rangle$ is not known as it usually does not arise in problems dealing with magnetism where the applied field is uniform. Combining these observation, we get

$$\langle \mathbf{R}_1^2 \mathbf{R}_2^2 \rangle \;=\; \sum_{ij=k_1}^{k_2} \sum_{kl=k_3}^{k_4} \left(6 \langle r_i^x r_j^x r_k^z r_l^z \rangle + 3 \langle r_i^z r_j^z r_k^z r_l^z \rangle \right) \quad . \tag{14}$$

In A, we derive the correlators $\langle r_i^x r_j^x r_k^z r_l^z \rangle$ and $\langle r_i^x r_j^z r_k^x r_l^z \rangle$ for $i \leq j \leq k \leq l$. The result reads

$$\langle r_i^x r_j^x r_k^z r_l^z \rangle \;=\; \langle r_i^z r_j^z r_k^x r_l^x \rangle = u^{j-i+l-k} \left(\frac{1}{15} v^{k-j} + w \frac{v^{k-j}-1}{v-1} \right) \quad , \tag{15}$$

$$\langle r_i^x r_j^z r_k^x r_l^z \rangle \;=\; \langle r_i^x r_j^z r_k^z r_l^x \rangle = \frac{1}{15} u^{j-i+l-k} t^{k-j} \quad . \tag{16}$$

We also recall here the result of [16] on $\langle r_i^z r_j^z r_k^z r_l^z \rangle$:

$$\langle r_i^z r_j^z r_k^z r_l^z \rangle = \frac{1}{9} u^{j-i} \left(\frac{4}{5} v^{k-j} + 1 \right) u^{l-k} \quad . \tag{17}$$

In the above formulae, we have introduced

$$v \;=\; 1 - 3\frac{u}{K} \quad , \quad w = \frac{1}{3} \left(K^{-1} \coth(K) - K^{-2} \right) \quad ,$$
$$t \;=\; \frac{1}{K^2} \left(3 + K^2 - 3K \coth(K) \right) \quad . \tag{18}$$

With these exact results, we can decompose the summations involved in (10) by symbolic computation [19]. The results for varying coupling constant K and chain overlap ζ are presented in the following section along with the results obtained from direct numerical computation.

B. Numerical results

c. ensemble averages We numerically evaluate the covariance using a Monte Carlo sampling of the classical one-dimensional Heisenberg model. A new trial configuration was obtained from the arbitrary reorientation of a randomly chosen spin. For small interaction strengths, such a sampling scheme achieves a high acceptance ratio. However, at stronger couplings, this ratio quickly drops. We evaluate the covariance $C(R_1^2, R_2^2)$ with $R_1 = \sum_{i=k_1}^{k_2} \mathbf{r}_i$ and $R_1 = \sum_{i=k_1}^{k_2} \mathbf{r}_i$, which we will denote from now on with the shorthand $C_{k_1 k_2 k_3 k_4}(K)$ to indicate its dependence on the coupling strength K. On the left hand side of figure 3, we show a comparison between the analytical and numerical results for Heisenberg model or, equivalently, the discrete flexible chain as a function of the coupling strength K for a small chain with $N = 5$ segments. We used 10^8 Monte Carlo steps for each value of the coupling constant, and chose $(k_1, k_2, k_3, k_4) = (1, 3, 2, 4)$, i.e. $\zeta = 2$. Notice that for $K = 0$, we recover the limit of the non-interacting freely-jointed chain, $C_{1324}(0)/b^4 = 4/3$. For small coupling K, the correction to the zero-coupling limit is linear, but the increase is sharper for larger K and passes through a maximim. Then, for large values of K, the covariance decreases with K towards zero. These results can be interpreted in view of the stiffening of the chain with increasing K: Whereas a large coupling constant initially tends to enhance cooperative motion among the segments, it also reduces the volume of the configurational space sampled on average as the energy penalty for overlapping monomers increases. This situation is schematically represented on the left hand side of figure 3, where snapshots of typical configurations of a dynamical implementation of the discrete flexible chain (discussed in the next subsection) for values of K below and above the maximum are shown. These can been understood more quantitatively by studying the joint probability distribution function $P(R_1^2, R_2^2)$ of the two distances. Figure 4 showns $P(R_1^2, R_2^2)$ in the previous case $(k_1, k_2, k_3, k_4) = (1, 3, 2, 4)$ for different values of the coupling constant K: at zero interaction in the freely-jointed limit (left-hand panel), at $K = K_{max} \approx 1.126$ which maximizes $C_{1324}(K)$ in figure 3 (middle panel), and at $K = 8$ (right-hand panel). One observes that the transition $C_{1324}(K) \rightarrow 0$ for large K goes
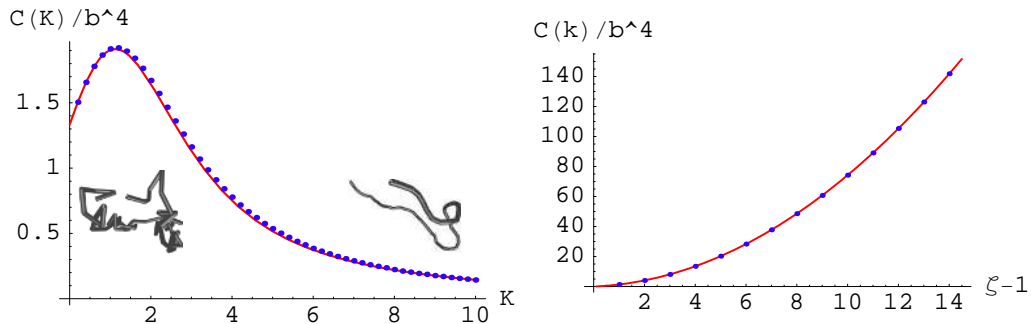
FIG. 3: Numerical comparison of the analytical result for the freely-jointed chain (solid line) with MC sampling of the discrete flexible chain as a function of interaction strength K; *Left:* covariance $C_{k_1 k_2 k_3 k_4}(K)$ for $(k_1, k_2, k_3, k_4) = (1, 3, 2, 4)$; points: results from MC sampling of a chain of length $N = 5$ with 10^8 steps; line: analytical results using (10),(12),(13),(15-17); *Right:* Covariance $C_{1k_2 k_3 (k_2+1)}(K)$ for K with varying overlap $\zeta = k_2 - k_3 + 1$ from MC sampling (points); the solid line shows the first order linear approximation $C_{1k_2 k_3 (k_2+1)}(K)/b^4 \approx \frac{2}{3}\zeta(\zeta - 1) + \frac{8}{9}K(\zeta - 1)^2$
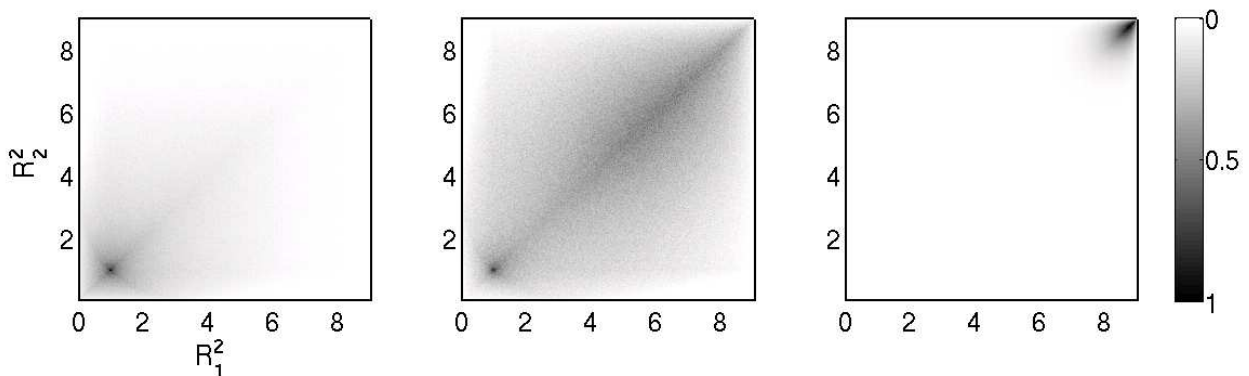

FIG. 4: Joint probability distribution function $P(R_1^2, R_2^2)$ for different coupling constants in the case $(k_1, k_2, k_3, k_4) = (1, 3, 2, 4)$ obtained from MC sampling as in figure 3. From left to right, the coupling constants are $K = 0$, $K = K_{max} \approx 1.126$, $K = 8$, gray and black colors indicate regions of higher probability (each panel normalized by $\max_{R1,R2} P(R_1^2, R_2^2)$).

along with a restriction to a relatively small part of configurational space associated to a stretched configuration of the polymer.

For small values of the coupling constant, the first order linear correction term can be estimated by performing simulations for varying overlap $\zeta = k_2 - k_3 + 1$. On the right hand side of figure 3, we evaluate the covariance function $C_{k_1 k_2 k_3 k_4}(K)$ for $K = 0.01$ fixed, and with $\zeta = 2, ..., 15$ along with a corresponding increase of the chain size L. The results are fitted by the first order linear approximation function $C_{1k_2 k_3 (k_2+1)}(K)/b^4 \approx \frac{2}{3}\zeta(\zeta - 1) + \frac{8}{9}K(\zeta - 1)^2$.

d. time averages Similarly to the non-interacting case, we would like to estimate the covariance from time averages using a dynamical implementation of the model. As outlined in section III B, we implement a distance constrained chain of mass points, while adding an additional contribution to the potential energy function $V_{KP} = -K \sum_{i=1}^{N-1} (\hat{\mathbf{r}}_i \cdot \hat{\mathbf{r}}_{i+1} - 1)$ that accounts for the energy penalty for monomer overlap in the discrete flexible chain. In the case $K = 0$, we recover the previous non-interacting case while for non-vanishing coupling values, the discrete flexible chain generates an orientation-dependent force. Here, we used a chain of $N = 7$ segments to estimate the running-time average

$$\mathcal{C}_{k_1 k_2 k_3 k_4}(K, T) = \overline{R_1^2 R_2^2} - \overline{R_1^2}\ \overline{R_2^2} \tag{19}$$

with $\overline{A(t)} = \frac{1}{T} \int_{t_0}^{t_0+T} dt'\ A(t')$ as defined previously. As an example, we chose $(k_1, k_2, k_3, k_4) = (1, 4, 2, 5)$. The simulation parameters were the same as described in section III B except for a smaller time-step $dt = 0.02$ and an equilibration time of $5 \cdot 10^3$ time units. Configurations were recorded for averaging every 20 time-steps. Figure 5 shows the evolution of the running-time average on a logarithmic scale for different values of the coupling constant ($K = 0.8, 1.6, 2.4, 3.6, 4.4, 5.2, 6, 8, 10$). Notice that these couplings include values below and above the coupling
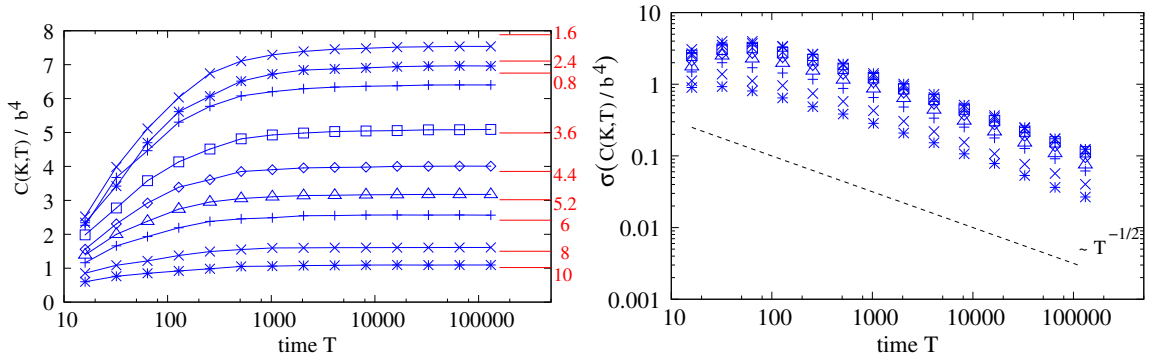
FIG. 5: *Left:* Time convergence of running time average of the covariance $\mathcal{C}_{k_1 k_2 k_3 k_4}(K, T)$ for different coupling constants K. Here, $(k_1, k_2, k_3, k_4) = (1, 4, 2, 5)$ and $N = 7$; the blue lines and points show the running time estimation of the covariance averaged over 1000 independent initial conditions, the red lines indicate the exact result. *Right:* Time convergence of the standard deviation of the distribution of covariances from different initial conditions. In both figures, the unit of time is $20dt$ (see text).
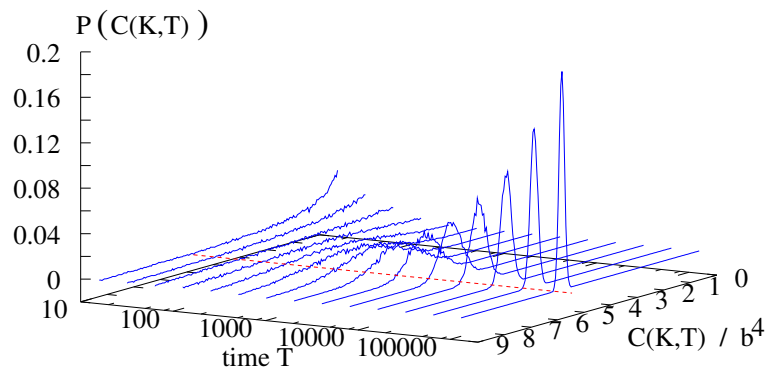

FIG. 6: Probability distribution function of the time averaged covariance function obtained from 10000 independent initial conditions, $(k_1, k_2, k_3, k_4) = (1, 5, 4, 6)$, $N = 7$ and $K = 0.4$. The red line indicated the analytically predicted value. The unit of time is $20dt$ (see text).

maximizing the covariance. For each K, the results were averaged over 1000 initial conditions to avoid statistical dependence of the results on a particular initial condition for the trajectory. While all simulations converge towards values close to the analytical results indicated by the red lines on the left hand side of figure 5, one observes that within the time of simulation this value is not exactly reached, and the relative changes towards long times are small. This can be seen more quantitatively by taking a closer look at the standard deviation of the covariance at different points in time. On the right hand side of figure 5, we computed the standard deviation σ of the covariance estimated from averages over different initial conditions. For long times, the variance decreases approximately as $T^{-1/2}$, indicating a slow change in time. For these timescales, increasing time is equivalent to increasing the number of statistically independent samples, such that the overall convergence behavior is Gaussian. This convergence behaviour is not limited to this particular observable, but can also obtained for other variables such as $\overline{R_1^2}$. An illustration of the convergence of this distribution is given in figure 6, where the probability distribution function of the covariance is shown as a function of time based on an estimation from 10000 initial conditions. The scale of binning of the histograms at each time step is based on the large distribution of values at the smallest timescale. One observes that the onset of the scaling with $T^{-1/2}$ in figure 5 is reflected here by the evolution toward distributions with a well defined peak and decreasing width at half maximum for long times.

V. SUMMARY AND OUTLOOK

We introduced the covariance of distances as a simple measure allowing to characterize correlated and uncorrelated motion in polymers and biomolecules. In this work, we focussed on the exactly solvable cases only. We considered two highly simplified models of homogeneous polymers. These can be though of as some zeroth-order approximation

to the real situation. A result which may seem surprising at first sight is that even a system without interacting forces can display a non-vanishing covariance stemming merely from a geometric overlap in the distances. Moving to a more realistic system, the discrete flexible chain, we have shown that the results of the non-interacting chain can be recovered as a special case of the covariance for the system with interaction. Also, we have shown that for small values of the coupling constant, i.e. for small excluded-volume penalties or, equivalently, high temperatures, the correction to the results of the freely-jointed chain scales linear with the coupling constant. This effect is due to an increase of correlated motion. This trend is yet not monotonic. With increasing coupling constant, the covariance at small couplings reaches a maximum and then decreases to zero. As a consequence, there exists a finite value of K for which the ensemble average of the covariance coincides with the value obtained at zero interaction. A simple interpretation of this observation is that though stronger coupling increases the cooperativity, it also reduces the configurational space sampled as the penalty of the excluded volume increases. Consequently, the average amplitude of distance fluctuation decreases with K. In section IV, we also analyzed the convergence of the time average of the covariance measure towards the ensemble average obtained from an exact calculation. While the quantitative difference between the predicted value and the results from numerical experiments certainly depend on the coupling to the stochastic heat bath, the qualitative results indicating that an algebraic decay in the variance would in principle occur for any type of canonical sampling provided the time scale is long enough so as to decorrelate events along individual trajectories. Regarding the situation in single molecule experiments, though our results have been in a highly idealized setting, some notable conclusions can be drawn. In single molecule FRET experiments, the acquisition time of the intensity signal is always limited either by instrumental factors or the destruction of the fluorescent marker by photobleaching. Accordingly, one can only expect to find meaningful results if the data acquisition time is much longer than the slowest timescale of the underlying dynamics of the molecule, and if the experiment can be sufficiently repeated so as to allow one for averages over independent trajectories. In the present model where the interaction is dominated by a single parameter, it appears that the approach to convergence is not affected by the strength of interaction. Though the expectation value of the covariance varies with K, the approach towards this value based on running time averages is K-independent. The convergence behavior in this model is not sufficient to draw conclusions on the physical state of the system (strongly/weakly interacting), but the mean value of the covariance is related (figure 3 in section IV). It therefore appears compulsory for an experiment aiming to distinguish among different states to be calibrated on an absolute intensity scale. From the results of the idealized models considered in this work, a promising future direction appears to be to apply the covariance analysis to more complex models of biopolymers where analytical approaches are not feasible anymore. Even if the ensemble average cannot be evaluated exactly, numerical experiments characterizing the dynamical convergence, similar to the ones performed on the polymer models in this article, are still feasible on long timescales. As an example, for proteins, a natural question arises whether the convergence of the covariance of distances within secondary structure elements differs significantly form the relaxation timescales on the level of tertiary structure. In particular, it would be promising to look at these cases under different physical conditions related to the folding process. The results of such a study might provide insight on how to choose, from a very large number of possibilities, particularly interesting locations for fluorescent labelling in single molecule spectroscopy, and hence getting a closer view on "molecules at work" beyond autocorrelation functions.

APPENDIX A: DERIVATION OF 4-TH ORDER ELEMENTS FOR THE DISCRETE FLEXIBLE CHAIN

In what follows, the derivation of the 4-th order elements required for the evaluation of the covariances in section IV is outlined in a more general way than required for the purpose of the present work by considering a chain with non-homogeneous couplings.

1. The set of angles for integration

We choose the following coordinate system for integration (see illustration 7). All unitary vectors $\mathbf{r}_\ell$ are marked by their polar angles with respect to a fixed frame (Θ_ℓ, Φ_ℓ). These are the so-called laboratory angles. To simplify the notation, we adopt the definition $|\mathbf{r}_j| = 1$ omitting the hat. We also choose a reference vector $\mathbf{r}_i$ labelled by its

laboratory angles. All other vectors $\mathbf{r}_j$ are marked by their polar coordinates (θ_j, φ_j) with respect to the axial vector $\mathbf{r}_{j-1}$ if $j > i$ and $\mathbf{r}_{j+1}$ if $j > i$. Then for $j > i$, the vector $\mathbf{r}_{j+1}$ is deduced from $\mathbf{r}_j$ by a rotation of $\mathbf{r}_j$ of axis $\mathbf{z} \wedge \mathbf{r}_j$ and angle θ_{j+1} followed by a rotation of axis $\mathbf{r}_j$ and angle φ_{j+1}. The situation is analogous for $j < i$. A rotation of angle θ and around a unitary axis $\mathbf{a}$ acts on a vector $\mathbf{x}$ according to

$$R_{\mathbf{a},\theta} \cdot \mathbf{x} = \cos(\theta) \mathbf{x} + (1 - \cos(\theta))(\mathbf{a} \cdot \mathbf{x}) \mathbf{a} + \sin(\theta)(\mathbf{a} \wedge \mathbf{x}) . \tag{A1}$$

An association of the both aforementioned rotations gives

$$\mathbf{r}_{j+1} = (\cos(\Theta_{j+1}) + \sin(\theta_{j+1}) \cot(\Theta_j) \cos(\varphi_{j+1})) \mathbf{r}_j - \sin(\theta_{j+1}) \begin{pmatrix} \sin(\Phi_j) \sin(\varphi_{j+1}) \\ \cos(\Phi_j) \sin(\varphi_{j+1}) \\ \cos(\varphi_{j+1}) / \sin(\Theta_j) \end{pmatrix} . \tag{A2}$$

(A2) constitutes a recursion relation for the laboratory angles (Θ_j, Φ_j) in terms of those located lower in the chain and of the local angles (θ_j, φ_j). The longitudinal recurrence is simple

$$\cos(\Theta_{j+1}) = \cos(\theta_{j+1}) \cos(\Theta_j) - \sin(\theta_{j+1}) \cos(\varphi_{j+1}) \sin(\Theta_j) , \tag{A3}$$

whereas the transverse one is a little more involved:

$$\mathbf{r}_{j+1}^{\perp} = \cos(\theta_{j+1}) \mathbf{r}_j^{\perp} + \sin(\theta_{j+1}) \begin{pmatrix} \cos(\varphi_{j+1}) \cos(\Phi_j) \cos(\Theta_j) - \sin(\varphi_{j+1}) \sin(\Phi_j) \\ \cos(\varphi_{j+1}) \sin(\Phi_j) \cos(\Theta_j) + \sin(\varphi_{j+1}) \cos(\Phi_j) \end{pmatrix} . \tag{A4}$$

These information are enough to compute the two and four-point correlation functions. The latter will be computed for an inhomogeneous Hamiltonian

$$H = -k_B T \sum_{j=1}^{N-1} b_{j+1} \mathbf{r}_j \cdot \mathbf{r}_{j+1} \qquad |\mathbf{r}_j| = 1 . \tag{A5}$$

The results of interest will follow either by taking the homogeneous limit, what we do in section IV where $b_i = K$, or by assuming that the couplings $\{b_i\}$ are statistical independent variables, and taking the ensemble average over such a distribution. We start by re-deriving the known results for the two-point functions as the techniques involved there will be used for the higher-order correlators.

2. The two-point function

Let $d\omega = \sin(\theta) d\theta d\varphi$ and $d\Omega = \sin(\Theta) d\Theta d\Phi$. One has

$$\int d\omega e^{b \cos(\theta)} = 4\pi \mathrm{shc}(b) \equiv 4\pi \frac{\sinh(b)}{b} . \tag{A6}$$

Then, for $i \leq j$, one has for the only non-trivial correlator

$$\langle r_i^z r_j^z \rangle = \int \frac{d\omega_1}{4\pi \mathrm{shc}(b_2)} \cdots \frac{d\omega_{i-1}}{4\pi \mathrm{shc}(b_i)} \frac{d\Omega_i}{4\pi} \frac{d\omega_{i+1}}{4\pi \mathrm{shc}(b_{i+1})} \cdots \frac{d\omega_N}{4\pi \mathrm{shc}(b_N)} r_i^z r_j^z \exp\left\{ \sum_{k=1}^{i-1} b_{k+1} \cos(\theta_k) + \sum_{k=i+1}^{N} b_k \cos(\theta_k) \right\}$$

$$= \int \frac{d\Omega_i}{4\pi} \frac{d\omega_{i+1}}{4\pi \mathrm{shc}(b_{i+1})} \cdots \frac{d\omega_j}{4\pi \mathrm{shc}(b_j)} e^{\sum_{k=i+1}^{j} b_k \cos(\theta_k)} r_i^z r_j^z . \tag{A7}$$

After this first trivial integration, in order to deal with the coupled integrals, we use the recursion of the angles. As Θ_{j-1} is independent of (θ_j, φ_j) we can integrate on the latter variables. We set

$$u(b) = \int \frac{d\omega}{4\pi \mathrm{shc}(b)} \cos(\theta) e^{b \cos(\theta)} = \frac{4\pi}{b^2} \frac{b \cosh b - \sinh b}{\mathrm{shc}(b)} , \tag{A8}$$

and since $\int \mathrm{d}\omega \sin(\theta)\cos(\varphi)\,\mathrm{e}^{b\cos(\theta)} = 0$, we get

$$\langle r_i^z r_j^z \rangle = u\,(b_j)\,\langle r_i^z r_{j-1}^z \rangle = \prod_{k=i+1}^{j} u\,(b_k) \int \frac{\mathrm{d}\Omega}{4\pi}\cos^2(\Theta) = \frac{1}{3}\prod_{k=i+1}^{j} u\,(b_k) \tag{A9}$$

recovering the result of Fisher[17]. A similar calculation can be done for the four-point functions. Here and in the following, we assume that $i \le j \le k \le \ell$. Due to a trivial integration on the angles located before the ith site, one has

$$\langle r_i^{a_i} r_j^{a_j} r_k^{a_k} r_\ell^{a_\ell} \rangle_b = \langle r_1^{\widetilde{a_1}} r_{\widetilde{j}}^{\widetilde{a_j}} r_{\widetilde{k}}^{\widetilde{a_k}} r_{\widetilde{\ell}}^{\widetilde{a_\ell}} \rangle_{\widetilde{b}} \quad . \tag{A10}$$

Where $\widetilde{j} = j - i + 1$ and $\widetilde{b_k} = b_{i+k-1}$, $\widetilde{a_k} = a_{i+k-1}$. Hence, we can compute the correlators starting from the first site and then, if needed, perform the aforementioned shifts.

We start with the XXZZ case. There, we choose $\mathbf{r}_j$ to be the reference vector for all of the "moving frame" angles. Thus

$$\langle r_1^x r_j^x r_k^z r_\ell^z \rangle = \int \prod_{p=1}^{j-1} \frac{\mathrm{d}\omega_p \mathrm{e}^{b_{p+1}\cos(\theta_p)}}{4\pi\,\mathrm{shc}\,(b_{p+1})} \frac{\mathrm{d}\Omega_j}{4\pi} \prod_{p=j+1}^{\ell} \frac{\mathrm{d}\omega_p \mathrm{e}^{b_p\cos(\theta_p)}}{4\pi\,\mathrm{shc}\,(b_p)} r_1^x r_j^x r_k^z r_\ell^z \quad . \tag{A11}$$

The recursion relations for the angles read

$$\begin{align}
r_p^x &= \cos(\theta_p)\,r_{p+1}^x + f_x\,(\cos(\varphi_p),\sin(\varphi_p)) && p \in [\![\,1\,;\,j-1\,]\!]\,, \tag{A12}\\
r_p^z &= \cos(\theta_p)\,r_{p-1}^z + f_z\,(\cos(\varphi_p),\sin(\varphi_p)) && p \in [\![\,k+1\,;\,\ell\,]\!]\,. \tag{A13}
\end{align}$$

The functions $f_{x,z}$ are linear in $\cos(\varphi_p)$ and $\sin(\varphi_p)$. Thence, their integrals over the azimuthal angle φ_p give zero. We are thus led to the same recursion as for the two-point function. Eventually,

$$\langle r_1^x r_j^x r_k^z r_\ell^z \rangle = \prod_{p=2}^{j} u\,(b_p) \prod_{p=k+1}^{\ell} u\,(b_p) \int \frac{\mathrm{d}\Omega_j}{4\pi} \left(r_j^x\right)^2 \left(r_k^z\right)^2 \prod_{p=j+1}^{k} \frac{\mathrm{d}\omega_p \mathrm{e}^{b_p\cos(\theta_p)}}{4\pi\,\mathrm{shc}\,(b_p)} \,. \tag{A14}$$

The recurrence equation for the squared laboratory angles reads:

$$\left(r_k^z\right)^2 = \left(r_{k-1}^z\right)^2 \left(\cos^2\theta_k - \sin^2\theta_k\cos^2\varphi_k\right) + \sin^2\theta_k\cos^2\varphi_k - \frac{1}{2}\sin 2\theta_k\cos\varphi_k\sin 2\Theta_{k-1}\,. \tag{A15}$$

The linear term in $\cos\varphi_k$ can be dropped after an integration over the azimuthal angle, so that after agreeing upon

$$v\,(b) = \int \frac{\mathrm{d}\omega}{4\pi\,\mathrm{shc}\,(b)}\mathrm{e}^{b\cos(\theta)}\left(\cos^2\theta - \sin^2\theta\cos^2\varphi\right) = 1 - \frac{3}{b^3}\left(\frac{b\cosh b - \sinh b}{\mathrm{shc}\,(b)}\right)\,, \tag{A16}$$

$$w\,(b) = \left(\int \frac{\mathrm{d}\Omega}{4\pi}\left(r^x\right)^2\right) \int \frac{\mathrm{d}\omega}{4\pi\,\mathrm{shc}\,(b)}\mathrm{e}^{b\cos(\theta)}\sin^2\theta\cos^2\varphi = \frac{b\cosh b - \sinh b}{3b^3\mathrm{shc}\,(b)}\,, \tag{A17}$$

we get

$$I_k = v\,(b_k)\,I_{k-1} + w\,(b_k) \quad \text{where} \quad I_k = \int \frac{\mathrm{d}\Omega_j}{4\pi}\left(r_j^x\right)^2\left(r_k^z\right)^2 \prod_{p=j+1}^{k} \frac{\mathrm{d}\omega_p \mathrm{e}^{b_p\cos(\theta_p)}}{4\pi\,\mathrm{shc}\,(b_p)}\,. \tag{A18}$$

This is a linear recursion whose homogeneous solution is $I_k = \prod_{p=j+1}^{k} v\,(b_k)\,I_j$. Setting $\widetilde{I}_k = I_k \prod_{p=j+1}^{k} v^{-1}\,(b_p)$ we get

$$\widetilde{I}_k - \widetilde{I}_{k-1} = \left(\frac{w\,(b_k)}{v\,(b_k)}\right) \prod_{p=j+1}^{k-1} v^{-1}\,(b_p)\,. \tag{A19}$$

Thus, the overall solution reads

$$I_k = \frac{1}{15} \prod_{p=j+1}^{k} v(b_k) + \sum_{p=j+1}^{k} w(b_p) \prod_{\ell=p+1}^{k} v(b_\ell) \ . \tag{A20}$$

where we have written the explicit value of I_j. This expression can be slightly simplified if one takes averages over the variables b. Assuming that these are independent random statistical variables, we get from averaging $[\cdot]$ over disorder

$$I_k = \frac{1}{15}[v(b)]^{k-j} + [w(b)] \sum_{p=j+1}^{k} [v(b)]^{k-p} = \frac{1}{15}[v(b)]^{k-j} + [w(b)]\frac{[v(b)]^{k-j} - 1}{[v(b)] - 1} \ . \tag{A21}$$

Note that the results for a homogeneous Hamiltonian follow from the distribution $\delta(b_i - K)$. After performing the appropriate shifts we get

$$\langle r_i^x r_j^x r_k^z r_\ell^z \rangle = [u(b)]^{j-i+\ell-k} \left\{ \frac{1}{15}[v(b)]^{k-j} + [w(b)]\frac{[v(b)]^{k-j} - 1}{[v(b)] - 1} \right\} \ . \tag{A22}$$

We now pass to the XZXZ and XZZX cases. The first reduction is identical with the previous case. Namely

$$\langle r_1^x r_j^z r_k^x r_\ell^z \rangle = \prod_{p=2}^{j} u(b_p) \prod_{p=k+1}^{\ell} u(b_p) \int \frac{\mathrm{d}\Omega_j}{4\pi} r_j^x r_j^z r_k^z r_\ell^x \prod_{p=j+1}^{k} \frac{\mathrm{d}\omega_p e^{b_p \cos(\theta_p)}}{4\pi \mathrm{shc}(b_p)} = \langle r_1^x r_j^z r_k^z r_\ell^x \rangle \ . \tag{A23}$$

These equalities are due to the fact that, for $a = x, z$, one has a relation of the type

$$r_\ell^a = \cos\theta_\ell r_{\ell-1}^a + f^a(\cos(\varphi_\ell), \sin(\varphi_\ell)) \quad \text{with} \quad \int_0^{2\pi} d\varphi_\ell \ f^a(\cos(\varphi_\ell), \sin(\varphi_\ell)) = 0 \ . \tag{A24}$$

Thus moving r_ℓ^z to the position k produces the same coefficient as moving r_ℓ^x to the position k. When writing a recursion for $r_k^z r_k^x$ it is enough to keep the terms that produce non-zero value after an integration on the azimuthal variables φ_k. Thus, up to integration vanishing terms,

$$r_k^z r_k^x = \frac{r_{k-1}^z r_{k-1}^x}{2} \left(3\cos^2\theta_k - 1\right) \ . \tag{A25}$$

Setting

$$t(b) = \frac{1}{2} \int \frac{\mathrm{d}\omega}{4\pi \mathrm{shc}(b)} e^{b\cos(\theta)} (3\cos(\theta) - 1) = \frac{1}{b^2}\left(b^2 + 3 - 3\coth b\right) \ , \tag{A26}$$

we obtain

$$\langle r_1^x r_j^z r_k^x r_\ell^z \rangle = \frac{1}{15} \prod_{p=2}^{j} u(b_p) \prod_{p=k+1}^{\ell} u(b_p) \prod_{p=j+1}^{k} t(b_p) \ . \tag{A27}$$

Thus performing the ensemble average as well as the shifts, we get

$$\langle r_i^x r_j^z r_k^x r_\ell^z \rangle = \frac{1}{15}[u(b)]^{j-i+\ell-k}[t(b)]^{k-j} \ . \tag{A28}$$

REFERENCES
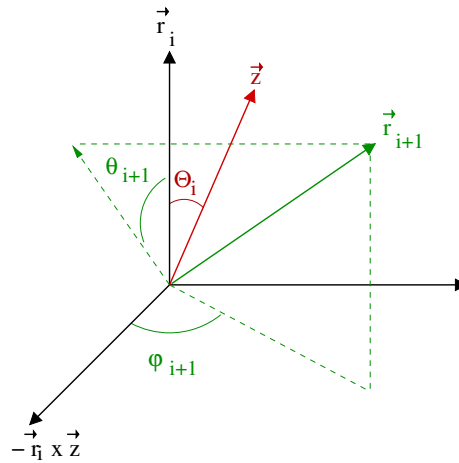
[1] Roy R, Hohng S and Ha T 2008 *Nature Methods* **5** 507

FIG. 7: illustration of the change of coordinates in (A2)

[2] Karplus M and McCammon J A 2002 *Nature Struc. Bio.* **9** 646
[3] van Gunsteren W F *et al.* 2006 *Angew. Chem. Int. Ed. Engl.* **45** 4064
[4] Brown W M, Martin S, Pollock S N, Coutsias E A and Watson J-P 2008 *J. Chem. Phys.* **129** 064118
[5] Yang H, Luo G, Karnchanaphanurach P, Louie T-M, Rech I, Cova S, Xun L and Xie X S 2003 *Science* **302** 262
[6] Hanson J A, Duderstadt K, Watkins L P, Bhattacharyya S, Brokaw J, Chu J-W and Yang H 2007 *Proc. Nat. Acad. Sci. USA* **104** 18055
[7] Hohng S, Joo C and Ha T 2004 *Biophys. J.* **87** 1328
[8] Watrob H M, Pan C-P and Barkley M D 2003 *J. Am. Chem. Soc.* **125** 7336
[9] Doi M and Edwards S F 1988 *The Theory of Polymers* (Oxford University Press)
[10] Izaguirre J A, Catarello D P, Wozniak J P and Skeel R D 2001 *J. Chem. Phys.* **114** 2090
[11] Allen M P and Tildesley D 1987 *Computer Simulations of Liquids* (Oxford, Clarendon Press)
[12] Franklin J and Doniach S 2006 *J. Chem. Phys.* **124** 154901
[13] Kratky O and Porod G 1949 *Recl. Trav. Chim. Pays Bas* **68** 1106
[14] Yan J, Kawamura R, and Marko J F 2005 *Phys. Rev. E* **71** 061905
[15] Mattis D C 1987 *The Theory of Magnetism* vol. I, ch. 6.5f (New York, Springer)
[16] Nakamura H and Takahashi M 1994 *J. Phys. Soc. Jpn* **63** 2563
[17] Fisher M E 1964 *Am. J. Phys.* **32** 343
[18] Tomita H and Mashiyama H 1972 *Prog. Theor. Phys.* **48** 1133
[19] Mathematica 5.0, Wolfram Inc.