Supplementary Material

Yen Ting (first draft of the derivations)

1 The generalized Brownian bridge

In this section, we outline the construction of the conditional process (the generalized "Brownian bridge") and the derivation of the mean and variance of such a conditional process.

We will first construct an underlying process (the random walk), which is assumed to be Gaussian and continuous in space. We release the constraint of discrete bead-and-rod model first, and use a continuous variable $s \in (0, Na)$ to denote the "coordinate" on the polymer. We remind the reader that N is the number of the bead, and a is the rod length. In real space, the location of the polymer at index $i \in \{1 \ldots N\}$ will then be denoted by $r(ia) \in \mathbb{R}^d$ in $d-$dimension.

The idea is to use this generalized Gaussian process to model the conformation of the bead-and-rod model. By "modeling", we mean that the Gaussian random walk respects the thermodynamical properties of the bead-and-rod model at the grid points $s = ia$, $i \in \{1 \ldots N\}$. Since the underlying process is Gaussian, we only need to specify the first two moments of the process (i.e. the mean $\langle r(s) \rangle$ and the covariance $\mathrm{cov}[r(s), r(t)]$ for any $s, t \in (0, Na)$).

Here, we emphasize the fact that the Gaussian random walk serves as a proxy

to represent the statistical properties of the bead-and-rod model on the grid $s = ia$, $i \in \{1 \ldots N\}$. We will adopt a heuristic approach to construct the process, and mathematically proving the rigor of their asymptotic equivalence at the lattice points is beyond our interests. At the end, we resort to numerical simulation (Monte Carlo in 1D and molecular dynamics in 3D) to show that in biologically relevant parameter regimes, the Gaussian process random walk approximates the bead-and-rod model very well.

In practice, we first construct the mean $m(s) := \langle r(s)|r(0) = 0 \rangle$ and the variance $\sigma_{k \to l}^2 := \mathrm{var}\,[r(la) - r(ka)]$ of a *unconditional* Gaussian random walk (see sections 2 and 3. The process is unconditional in the sense that we did not enforce the loop condition $r(Na) = 0 = r(0)$ first.) Assuming the mean $m(s)$ and any 2-point variance $\sigma_{s/a \to t/a}^2$ are given for any arbitrary $s, t \in (0, aN)$, the probability density function at location s is prescribed by

$$\rho_0\left\{r(s) = r_s\right\} = \rho(r(s) = r_s | r(0) = 0) = \frac{1}{\sqrt{2\pi\sigma_{0 \to s/a}^2}} \exp\left\{-\frac{[r_s - m(s)]^2}{2\sigma_{0 \to s/a}^2}\right\}. \quad (1)$$

We will always impose the "initial condition" $r(0) = 0$, and as a short-hand notation we shall use $\rho_0\{\cdot\}$ to represent $\rho\{\cdot|r(0) = 0\}$. Similar notation will be adopted for variance var_0 and covariance cov_0. In order to impose the loop condition, we

notice

$$\rho_0\left\{r(s) = r_s | r(aN) = 0\right\} = \frac{\rho_0\left\{r(s) = r_s, r(aN) = 0\right\}}{\rho_0\left\{r(aN) = 0\right\}} \tag{2a}$$

$$= \frac{\rho_0\left\{r(s) = r_s\right\}\rho\left\{r(s) = r_s, r(aN) = 0\right\}}{\rho_0\left\{r(aN) = 0\right\}} \tag{2b}$$

$$= \frac{\rho_0\left\{r(s) = r_s\right\}\rho\left\{r(0) = 0, r(aN - s) = r_s\right\}}{\rho_0\left\{r(aN) = 0\right\}} \tag{2c}$$

$$= \frac{\rho_0\left\{r(s) = r_s\right\}\rho_0\left\{r(aN - s) = r_s\right\}}{\rho_0\left\{r(aN) = 0\right\}} \tag{2d}$$

We established the second equation (2b) by the Markov property of the process. The symmetry under the transformation $s \to aN - s$ (i.e. indexing the polymer in the oposite direction) ensures (2c). Now, we insert (1) into (2d) and obtain

$$\rho_0\left\{r(s) = r_s | r(aN) = 0\right\} = \frac{\rho_0\left\{r(s) = r_s, r(aN) = 0\right\}}{\rho_0\left\{r(aN) = 0\right\}} \tag{3a}$$

$$\propto \frac{\exp\left(-\frac{[r_s - m(s)]^2}{2\sigma_{0\to s/a}^2}\right)\exp\left(-\frac{[r_s - m(aN - s)]^2}{2\sigma_{0\to N - s/a}^2}\right)}{\exp\left(-\frac{[r(aN) - m(aN)]^2}{2\sigma_{0\to N}^2}\right)}. \tag{3b}$$

Note that $m(s) = m(aN - s)$ since the symmetry of the problem must be preserved, and as a consequence

$$\rho_0\left\{r(s) = r_s | r(aN) = 0\right\} \propto \exp\left[-\frac{[r_s - m(s)]^2}{2}\left(\frac{1}{\sigma_{0\to s/a}^2} + \frac{1}{\sigma_{0\to N - s/a}^2}\right)\right], \tag{4}$$

which is again a Gaussian distribution with a mean $m(s)$ and a variance equal to the harmonic mean of $\sigma_{0\to s/a}^2$ and $\sigma_{0\to N - s/a}^2$ at s. In addition, we shall see that the sum $\sigma_{0\to s/a}^2 + \sigma_{0\to N - s/a}^2 = \sigma_{0\to N}^2$ by construction (see section 2 and 3), hence the

variance of the conditional process is

$$\sigma^2_{cond}(s) = \frac{\sigma^2_{0\to s/a}\sigma^2_{0\to N-s/a}}{\sigma^2_{0\to N}}. \tag{5}$$

Finally, we remark three facts:

- For any $s, t \in (0, aN)$, a similar derivation can be performed to derive the conditional covariance $\text{cov}_0[r(s), r(t)|r(aN) = 0]$ of the conditional Gaussian random walk and shows

$$\text{cov}_0\left[r(s), r(t)|r(aN) = 0\right] = \frac{\sigma^2_{0\to\min\{s,t\}/a}\sigma^2_{\max\{s,t\}/a\to N}}{\sigma^2_{0\to N}} \tag{6}$$

- We present a more concise form of the conditional Gaussian process by using the conventional notations in analysis of stochastic processes. The conditional process R_s^{cond} (here we adopt the conventional notation in stochastic processes that capital symbols represent the process) is

$$R_s^{cond} = R_s + \frac{\sigma^2_{0\to s/a}}{\sigma^2_{0\to N}}R_{aN} \tag{7}$$

where R_s is the unconditional process which is defined by

1. R_s is almost surely continuous,

2. $R_0 = 0$,

3. R_s has independent increments,

4. $R_t - R_s \sim \mathcal{N}(m(t) - m(s), \sigma^2_{s/a\to t/a})$, $\forall 0 < s \le t < aN$.

4

We introduce the process notation because it is easier[1] to perform calculations. In principle, one can calculate and prove any physical quantities of interests by either directly computing with the conditional probability distribution (4) and the conditional covariance (6), or alternatively solely by the corresponding process (7).

2 Constructing $m(s)$ and $\sigma^2_{k \to l}$ in 1D

In this section, we present a way to construct $m(s)$ and $\sigma^2_{k \to l}$ for the one dimensional model.

In principle, one completely solves problem after deriving the probabilities of

---

[1] For example, the conditional covariance can be calculated relatively easily (assume $0 < s < t < L := aN$):

$$\mathrm{cov}_0\left[r(s), r(t) | r(aN) = 0\right] = \mathrm{cov}_0\left[R_s^{cond}, R_t^{cond}\right]$$
$$= \mathrm{cov}_0\left[R_s - \alpha R_L, R_t - \beta R_L\right] \tag{8a}$$
$$= \mathrm{cov}_0\left[R_s, R_t\right] - \alpha\mathrm{cov}_0\left[R_s, R_L\right] \tag{8b}$$
$$- \beta\mathrm{cov}_0\left[R_L, R_t\right] + \alpha\beta\mathrm{cov}_0\left[R_L, R_L\right] \tag{8c}$$

where $\alpha := \sigma^2_{0 \to s/a}/\sigma^2_{0 \to N}$ and $\beta := \sigma^2_{0 \to t/a}/\sigma^2_{0 \to N}$. Each of the covariances can be computed relatively easily by the definition of independent increments; for example

$$\mathrm{cov}_0\left[R_s, R_L\right] = \mathrm{cov}_0\left[R_s, R_L - R_s + R_s\right]$$
$$= \mathrm{cov}_0\left[R_s, R_L - R_s\right] + \mathrm{cov}_0\left[R_s, R_s\right] = \mathrm{cov}_0\left[R_s, R_s\right] = \sigma^2_{0 \to s/a},$$

therefore

$$\mathrm{cov}_0\left[r(s), r(t) | r(aN) = 0\right] = \sigma^2_{0 \to s/a} - \alpha\sigma^2_{0 \to t/a} - \beta\sigma^2_{0 \to s/a} + \alpha\beta\sigma^2_{0 \to N}$$
$$= \sigma^2_{0 \to s/a} - \frac{\sigma^2_{0 \to s/a}\sigma^2_{0 \to t/a}}{\sigma^2_{0 \to N}} - \frac{\sigma^2_{0 \to s/a}\sigma^2_{0 \to t/a}}{\sigma^2_{0 \to N}} + \frac{\sigma^2_{0 \to s/a}\sigma^2_{0 \to t/a}}{\sigma^2_{0 \to N}}$$
$$= \frac{\sigma^2_{0 \to s/a}\sigma^2_{t/a \to N}}{\sigma^2_{0 \to N}}$$

which is exactly Eq.(6). In contrast, the computation by directly using the conditional probability distribution and covariances involves in two Gaussian integrals (at s and t) which is a straightforward but slightly more tedious calculation.

the occupation $\mathbb{P}\{n_j = 0\}$ for every single energy level j:

$$\mathbb{P}\{n_j = 1\} = \frac{1}{1 + \exp\left(\frac{j-\mu}{T}\right)}. \tag{11}$$

Here μ is the chemical potential and can be easily obtained with the aid of the geometrical symmetry ($\mathbb{P}\{n_j = 1\} = \mathbb{P}\{n_{N-j} = 0\} \ \forall j \in \{1 \ldots N\}$.) To obtain the unconditional mean and variance, one simply plug $r_i = a\left(2\sum_{j=1}^{i} n_j - i\right)$ back to the moments and obtain the mean

$$\langle r_k \rangle = a\left(2\sum_{l=1}^{k}\langle n_l \rangle - k\right), \tag{12}$$

and the variance

$$\mathrm{var}_0\left[r_l - r_k\right] = 4a^2\sum_{j=k+1}^{l}\mathrm{var}\left[n_j\right]. \tag{13}$$

Note that here we have used the assumption that n_j and n_k are independent if $j \neq k$, so that the variance of the sum (r_i) of the random variables (n_j) is exactly the sum of the variance of the random variables. *This is the only approximation.* When the number of particles N is large enough such an assumption is appropriate. The way to reintroduce the global dependence is to re-introduce an appropriate chemical potential μ at the end of the analysis. Such an approximation (by introducing independent distributions) is often implicitly introduced in most of the problems in condense matter physics or statistical mechanics, and we shall proceed our analysis with such an approximation.

We also notice that at this stage Eqs.(12) and (13) are only defined on $k, l \in \{1 \ldots N\}$ and insufficient to serve as the input to construct the effective Gaussian random walk R_s, which is a (almost surely) continuous process. However, since we

are only interested in the physical properties at on lattice $k, l \in \{1 \ldots N\}$, we can simply define two continuous functions $m(s)$ and $\sigma^2_{s/a \to t/a}$ such that their values on the lattice coincide with Eqs.(12) and (13):

$$m(s) \equiv \langle r_{s/a} \rangle, \text{ and} \tag{14a}$$

$$\sigma^2_{s/a \to t/a} \equiv \text{var}_0 \left[r_{s/a} - r_{t/a} \right], \ \forall s, t \in \{a, 2a, \ldots, Na\}. \tag{14b}$$

Philosophically, we try to extend Eqs.(12) and (13) by analytic continuation. Even though the extension is not unique, it is clear that any m and σ^2 functions satisfying Eqs.(14) predict the same statistical properties on lattice points—which are the physical quantities in which we are interested. In addition, note that by construction the variances are additive in the following sense:

$$\sigma^2_{k \to l} + \sigma^2_{l \to m} \equiv \sigma^2_{k \to m}. \tag{15}$$

In principle, we can now use Eqs.(14) to construct the *unconditional* Gaussian random walk (see previous Sec. 1) and we can easily obtain the mean and the variance of the *conditional* Gaussian random walk:

$$\mathbb{E} \left[R_i^{cond} a \right] = a \left(2 \sum_{j=1}^{i} \langle n_j \rangle - i \right), \tag{16a}$$

$$\text{var}_0 \left[R_i^{cond} a \right] = 4a^2 \frac{\left(\sum_{j=1}^{i} \text{var}\left[n_j\right] \right) \left(\sum_{k=i}^{N} \text{var}\left[n_j\right] \right)}{\sum_{l=1}^{N} \text{var}\left[n_j\right]}, \ \forall i \in \{1 \ldots N\}. \tag{16b}$$

The rest of the section devotes to analyze the asymptotic behavior of the solution (16) when the effective temperature is high $\tilde{T} \gg 1$. In this case, one can convert the sum into a integral by defining a continuous variable $\xi := j/\tilde{T}$ and

7

treat $1/\tilde{T}$ to be the small parameter, yielding

$$\sum_{j=1}^{i} \langle n_j \rangle = \sum_{j=1}^{i} \frac{1}{1 + \exp\left(\frac{j-\mu}{\tilde{T}}\right)} = \int_0^{iN} \frac{1}{1 + \exp\left(\xi - \frac{\mu}{\tilde{T}}\right)} \tilde{T} d\xi + \mathcal{O}(\tilde{T}^{-1}) \quad (17a)$$

$$= -\tilde{T} \log \frac{1 + \exp\left(\frac{-i+\mu}{\tilde{T}}\right)}{1 + \exp\left(\frac{\mu}{\tilde{T}}\right)} + \mathcal{O}(\tilde{T}^{-1}), \quad (17b)$$

therefore

$$\langle r_i \rangle = -2a\tilde{T} \left(\log \frac{1 + \exp\left(\frac{-i+\mu}{\tilde{T}}\right)}{1 + \exp\left(\frac{\mu}{\tilde{T}}\right)} - \frac{i}{2\tilde{T}} \right) + \mathcal{O}(\tilde{T}^{-1}) \quad (18a)$$

$$= 2a\tilde{T} \log \frac{1 + \exp\frac{\mu}{\tilde{T}}}{\exp\frac{i}{2\tilde{T}} + \exp\frac{2\mu-i}{\tilde{T}}} + \mathcal{O}(\tilde{T}^{-1}) \quad (18b)$$

Similarly, we can derive the high $\tilde{T}$ asymptote of the varraince by noticing the events $\{n_j = 1\}$ is a Bernoulli trial, so

$$\text{var}[n_j] = \mathbb{P}\{n_j = 1\}(1 - \mathbb{P}\{n_j = 1\}) = \frac{\exp\left(\frac{j-\mu}{\tilde{T}}\right)}{\left[1 + \exp\left(\frac{j-\mu}{\tilde{T}}\right)\right]^2}, \quad (19)$$

again by converting the sum to integral, we have

$$\sigma_{k\to l}^2 = 4a^2 \left[\frac{1}{1 + \exp\frac{k-\mu}{\tilde{T}}} - \frac{1}{1 + \exp\frac{l-\mu}{\tilde{T}}} \right] + \mathcal{O}(\tilde{T}^{-1}). \quad (20)$$

By plugging (20) back to (5), we obtain the high (effective) temperature asymptotic solution of the variance conditioning on the loop geometry. Finally, we use the symmetric condition $m(s) = m(aN - s)$ to determine μ has to be $N/2$ in this asymptotic formulation.

3 Deriving continuous asymptotes in 3D

The analysis of the 3D case is very similar to the 1D case but slightly more involved. Here we provide a outline of the analysis for completeness.

Model setting: each bond is treated as a spin $\vec{s} \in \mathbb{R}^3$ such that $|\vec{s}| = a$. Let the force is pointing to the $+z$ direction, and we take the normal parametrization of the spherical coordinate (ρ, θ, ϕ) such that the projection of the spin onto $\hat{z}$ to be $\cos \theta$. The potential energy of the chain is

$$U = -\gamma a v_0 \sum_{i=1}^{N} \sum_{j-1}^{i} \cos \theta_j. \tag{21}$$

Now we exchange the order of the double sum which yields

$$U = \gamma a v_0 \sum_{j=1}^{N} cos\theta_j \left(N - j + 1\right) = U_0 + \gamma a v_0 \sum_{j=1}^{N} j \cos \theta_j. \tag{22}$$

The last equality comes from the loop constraint $\sum_{j=1}^{N} \cos \theta_j = 0$ since the polymer has to come back to the origin after N steps.

Then, the grand partition function is

$$\mathcal{P} = \prod_{j=1}^{N} \mathcal{P}_j = \prod_{j=1}^{N} \int_{0}^{2\pi} d\phi_j \int_{0}^{\pi} \sin \theta_j \ d\theta_j e^{-\frac{\gamma a v_0 (j-\mu) \cos(\theta)}{k_B T}}. \tag{23}$$

The chemical potential for the spin pointing to (θ_j, ϕ_j) has to be proportional to $\cos \theta_j$, as easily seen by differential free energy $dF \sim \mu_{\theta_j} dN_{\theta_j} \propto \cos \theta_j dN_{\theta_j}$. This (strong) condition suggests the existence of a global chemical potential μ, as seen in (23). Note that the grand partition function factorizes with our assumption of independence of the configurations of the spins i and j (if $i \neq j$).

After the integration, we have

$$\mathcal{P}_j \propto \frac{k_B T \sinh \frac{\mu - j}{k_B T}}{\gamma a v_0 \left(\mu - j\right)}, \tag{24}$$

which has a well-defined limit as $\mu \to j$ for any j. Note that $\mathcal{P}_j$ is an even function of $(\mu - j)$ and so $\partial_\mu \mathcal{P}_j$ is an odd function of $(\mu - j)$. Clearly the effective temperature of the 3D model is $2\tilde{T} = k_B T / (\gamma a v_0)$.

By standard technique in statistical mechanics, we can derive the mean and the variances of the spin configurations:

$$\langle \cos \theta_j \rangle = 2\tilde{T} \partial_\mu \mathcal{P}_j = \coth \frac{\mu - j}{2\tilde{T}} - \frac{2\tilde{T}}{\mu - j}, \tag{25a}$$

$$\langle \cos^2 \theta_j - \langle \cos \theta_j \rangle^2 \rangle = \left(2\tilde{T} \partial_\mu\right)^2 \mathcal{P}_j = \left(\frac{2\tilde{T}}{\mu - j}\right)^2 - \operatorname{csch}^2 \frac{\mu - j}{2\tilde{T}}, \tag{25b}$$

The "coming back home" condition requires

$$\sum_{j=1}^{N} \langle \cos \theta_j \rangle = 0. \tag{26}$$

Note that the function $\langle \cos \theta_j \rangle$ is odd in $j - \mu$, so μ must be set to be $(N+1)/2$ to satisfy (26).

In the directions of x and y, we first notice that the azimuthal symmetry immediately predicts the $\langle \hat{x} \cdot \vec{s} \rangle = \langle \hat{y} \cdot \vec{s} \rangle = 0$, leaving only the variances of $\hat{x} \cdot \vec{s}$ and $\hat{y} \cdot \vec{s}$ undetermined. From the symmetry, we also know that the variance in the direction of $\hat{x}$ must be equal to the variance in the direction of $\hat{y}$. Moreover,

10

since the length of the spin is always 1, we can establish the relation

$$\left\langle (\vec{s} \cdot \hat{x})^2 \right\rangle = \left\langle (\vec{s} \cdot \hat{y})^2 \right\rangle = 1 - \left\langle \cos^2 \theta_j \right\rangle. \tag{27a}$$

Therefore, Eq.(25) uniquely determines the statistical properties of the 3-dimensional conformation of the polymer.

That being said, we are ready to construct the effective Gaussian random walk, as described in Sec. 1 and similar to the construction in Sec. 2. It should be clear that the mean in each direction can be built and satisfies

$$m_x(s) \equiv 0 \tag{28a}$$

$$m_y(s) \equiv 0 \tag{28b}$$

$$m_z(s) \equiv \sum_{j=1}^{s/a} \left(\coth \frac{\mu - j}{2\tilde{T}} - \frac{2\tilde{T}}{\mu - j} \right), \tag{28c}$$

$\forall s \in \{a, 2a, \dots, Na\}$, and the variance in each direction can be built and satisfies[2]

---

[2]One can again perform large $\tilde{T}$ asymptotic expansion and turn the sum into integrals:

$$\mathbb{E}\left[X_t | X_0 = 0\right] = \mathbb{E}\left[Y_t | Y_0 = 0\right] = 0, \tag{29a}$$

$$\mathbb{E}\left[Z_t | Z_0 = 0\right] = 2\tilde{T} \left[\log \frac{t - \frac{N}{2}}{2\tilde{T} \sinh\left(\frac{t - \frac{N}{2}}{2\tilde{T}}\right)} \right], \tag{29b}$$

$$\sigma^2_{l \to k; z} := \text{var}\left[Z(la) - Z(ka)\right] = 2\tilde{T} \left[\coth \frac{s - \frac{N}{2}}{2\tilde{T}} - \frac{1}{s - \frac{N}{2}} \right]_{s=ka}^{lk} \tag{29c}$$

$$\sigma^2_{l \to k; x} := \text{var}\left[X(la) - X(ka)\right] = \sigma^2_{l \to k; y} := \text{var}\left[Y(la) - Y(ka)\right]$$

$$= \int_{l/2\tilde{T}}^{k/2\tilde{T}} \left\{ 1 + \text{csch}^2\left(\frac{N}{4\tilde{T}} - \xi\right) - \left(\frac{1}{\frac{N}{4\tilde{T}} - \xi}\right)^2 \right.$$

$$\left. - \left[\coth\left(\frac{N}{4\tilde{T}} - \xi\right) - \left(\frac{1}{\frac{N}{4\tilde{T}} - \xi}\right) \right]^2 \right\} d\xi. \tag{29d}$$

Note that $\mu = N/2$ in this case. Since (29d) has no closed form to our knowledge and must

11

$$\sigma^2_{s/a \to t/a;x} = \sigma^2_{s/a \to t/a;y} \tag{30a}$$

$$\equiv \sum_{j=s/a}^{t/a} \left[1 - \frac{4\tilde{T}^2}{(\mu - j)^2} + \operatorname{csch}^2 \frac{\mu - j}{2\tilde{T}} + \left(\coth \frac{\mu - j}{2\tilde{T}} - \frac{2\tilde{T}}{\mu - j} \right)^2 \right] \tag{30b}$$

$$\sigma^2_{s/a \to t/a;z} \equiv \sum_{j=s/a}^{t/a} \left[\frac{4\tilde{T}^2}{(\mu - j)^2} - \operatorname{csch}^2 \frac{\mu - j}{2\tilde{T}} \right], \quad \forall s, t \in \{a, 2a, \dots, Na\}. \tag{30c}$$

Finally, we impose the loop condition to obtain the "generalized Brownian bridge":

$$X_s^{cond} = X_s - \frac{\sigma^2_{0 \to s/a;x}}{\sigma^2_{0 \to N;x}} X_{aN}, \tag{31a}$$

$$Y_s^{cond} = Y_s - \frac{\sigma^2_{0 \to s/a;y}}{\sigma^2_{0 \to N;y}} Y_{aN}, \tag{31b}$$

$$Z_s^{cond} = Z_s - \frac{\sigma^2_{0 \to s/a;z}}{\sigma^2_{0 \to N;z}} Z_{aN} \tag{31c}$$

where the unconditional processes X_s, Y_s, and Z_s are three independent and unconditional Gaussian random walk with their respective mean (28) and 2-point variance function (30).

Calculations of the statistical properties of the polymer loop under soft and hard constraints naturally follows by applying parallel analyses for 1D system in each of the direction x, y, and z. For example, it should be clear now that the variance of the distance between homologous sites of two independent and identically distributed (i.i.d.) chromosomes is equal to the sum of the variances of

_____

be evaluated numerically, we in stead numerically evaluate the discrete sum of the more general solutions (30) as our theoretical prediction for the 3D model.

the distances between homologous sites in each dimension,

$$\text{var}\left[\left|\vec{R}_s^{cond,(1)} - \vec{R}_s^{cond,(2)}\right|\right] = \sum_{R \in \{X,Y,Z\}} \text{var}\left[\left|R_s^{cond,(1)} - R_s^{cond,(2)}\right|\right] . \tag{32}$$

Furthermore, each of the variances in one specific dimension will be equal to twice of the variance of one independent chromosome at that point in that direction by the i.i.d. property,

$$\text{var}\left[\left|R_s^{cond,(1)} - R_s^{cond,(2)}\right|\right] = \text{var}\left[R_s^{cond,(1)}\right] + \text{var}\left[R_s^{cond,(2)}\right] = 2\text{var}\left[R_s^{cond}\right] . \tag{33}$$

4   Soft constraints: the presence of the centromere

The presence of the centromere imposes a "soft constraint" on a pair of chromosomes: the homologous sites are "glued" together at the location of the centromere, but unlike the spindle-pole body $(s = 0, aN)$ which is dragged at almost a constant speed (and hence pinned in space in the co-moving frame, $r(s) = r(aN) = 0$), the centromere is free to move in the space.

This section presents the derivation of the *variance of the distance* between a pair of the homologous sites of *two chromosomes* with the presence of a centromere:

$$\text{var}\left[\Delta R^{cond}\right] := \text{var}\left[R_s^{cond;(1)} - R_s^{cond;(2)} | R_t^{cond;1} = R_t^{cond;2}\right] . \tag{34}$$

The calculation presented is performed in one-dimensional space, since we have demonstrated the construction of three independent one-dimensional generalized Brownian bridges to characterize the conformation of chromosomes in 3-space in Sec. 3.
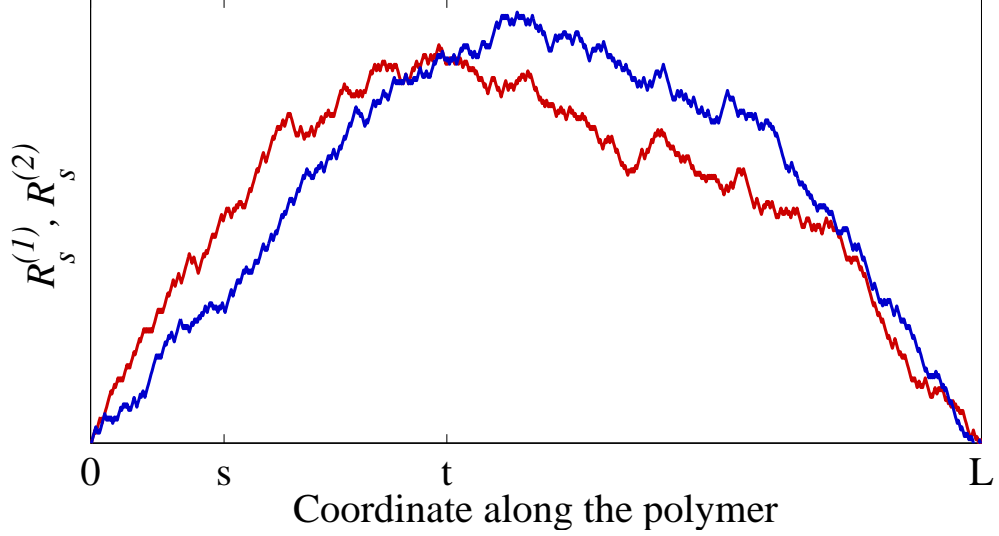
Figure 1: Illustration of two polymers constrained by a centromere at t.

We refer the readers to Fig. 1 to visualize the one dimensional problem: the blue curve and red curve represent a paired chromosomes. This pair of chromosomes are constrained by the presence of the centromere at polymer index t, *in addition to* the obvious geometrical (hard) constraint $R_L^{cond;(1)} = R_L^{cond;(2)} = R_0^{cond;(1)} = R_0^{cond;(2)} = 0$.

There are two ways to calculate the variance $\text{var}\left[\Delta R^{cond}\right]$. One way is to directly compute with conditional probability density from the propagating kernel (1) of two independent polymers, conditioning on the presence of *both* the soft and hard constraints. The second way is to perform calculation with the already derived conditional process (7). In a moment, we shall see the strength of the second approach.

We outline the calculation of the first approach. It is clear that the hard-

constrained polymers have the probability density at s:

$$\rho \left\{ r^{(1)}(s) = r_1, r^{(1)}(t) = r_C | r^{(1)}(0) = r^{(1)}(L) = 0 \right\}$$

$$\propto \exp \left[-\frac{r_1^2}{2\sigma_{0 \to s/a}^2} - \frac{(r_1 - r_C)^2}{2\sigma_{s/a \to t/a}^2} - \frac{r_C^2}{2\sigma_{t/a \to N}^2} \right], \qquad (35a)$$

$$\rho \left\{ r^{(2)}(s) = r_2, r^{(2)}(t) = r_C | r^{(2)}(0) = r^{(2)}(L) = 0 \right\}$$

$$\propto \exp \left[-\frac{r_2^2}{2\sigma_{0 \to s/a}^2} - \frac{(r_2 - r_C)^2}{2\sigma_{s/a \to t/a}^2} - \frac{r_C^2}{2\sigma_{t/a \to N}^2} \right], \qquad (35b)$$

$$\rho \left\{ r^{(1)}(t) = r^{(2)}(t) = r_C | r^{(1)}(0) = r^{(2)}(0) = r^{(1)}(N) = r^{(2)}(N) = 0 \right\}$$

$$\propto \exp \left[-2\frac{r_C^2}{2\sigma_{0 \to t/a}^2} - 2\frac{r_C^2}{2\sigma_{t/a \to N}^2} \right], \qquad (35c)$$

where we denote the position of the centromere to be r_C. As a consequence, the probability distribution with the soft constraint can be computed accordingly

$$\rho_s := \rho \left\{ r^{(1)}(s) = r_1, r^{(2)}(s) = r_2 | r^{(i)}(0) = r^{(i)}(L) = 0, r^{(i)}(t) = r_C, i \in \{1, 2\} \right\}$$

$$= \frac{\prod_{i \in \{1,2\}} \rho \left\{ r^{(i)}(s) = r_i, r^{(i)}(t) = r_C | r^{(i)}(0) = r^{(i)}(L) = 0 \right\}}{\rho \left\{ r^{(1)}(t) = r^{(2)} = r_C | r^{(i)}(0) = r^{(i)}(L) = 0, i \in \{1, 2\} \right\}}$$

$$(36a)$$

$$\propto \exp \left[-\frac{r_1^2 + r_2^2}{2\sigma_{0 \to s/a}^2} - \frac{(r_1 - r_C)^2 + (r_2 - r_C)^2}{2\sigma_{s/a \to t/a}^2} + 2\frac{y_C^2}{2\sigma_{0 \to t/a}^2} \right] \quad (36b)$$

Note that the probability density ρ_s depends on the position of the homologous sites r_1 and r_2, as well as the position of the centromere r_C. Since r_C can be any value in $\mathbb{R}^1$ (reflecting the fact that the constraint is "soft"), we will have to integrate over all the possibilities with the appropriate weight to obtain the final

probability density. Mathematically,

$$
\rho \left\{ r^{(1)}(s) = r_1, r^{(2)}(s) = r_2 | r^{(i)}(0) = r^{(i)}(L) = 0 \right\} = \int_{-\infty}^{\infty} \rho_s(r_1, r_2, r_C) dy_C
$$

$$
\propto \exp \left[-\frac{(r_1 - r_2)^2}{4 \frac{\sigma_{0 \to s/a}^2 \sigma_{0 \to t/a}^2}{\sigma_{0 \to t/a}^2}} \right]. \qquad (37a)
$$

From the expression of the probability distribution, we read off the variance of the distance between the homologous site

$$
\mathrm{var} \left[r^{(1)}(s) - r^{(2)}(s) \right] = 2 \frac{\sigma_{0 \to s/a}^2 \sigma_{0 \to t/a}^2}{\sigma_{0 \to t/a}^2}. \qquad (38)
$$

We now turn to the process formulation and demonstrate a much more concise derivation. As described in the text, we will relabel the chromosomes and build an effective process starting from spindle-pole body to the centromere along chromosome 1, and then progress "backward" along chromosome 2 back to the spindle-pole body again. The illustrative Fig. 2 is presented as a comparison to the joint process Fig. 1.

The unconditional effective process $\tilde{R}_{\mathrm{eff};s}$ for $s \in (0, L' := 2t)$ characterizes the above-mentioned relabeled process and satisfies

$$
\mathbb{E} \left[\tilde{R}_{\mathrm{eff};s} \right] = 0, \qquad (39a)
$$

$$
\mathrm{var}_0 \left[\tilde{R}_{\mathrm{eff};s} \right] = \sigma_{\mathrm{eff};0 \to s/a}^2 := \begin{cases} \sigma_{0 \to s/a}^2, & \text{if } s \leq t, \\ \sigma_{0 \to t/a}^2 + \sigma_{(t-s)/a \to t/a}^2, & \text{if } s > t, \end{cases} \qquad (39b)
$$

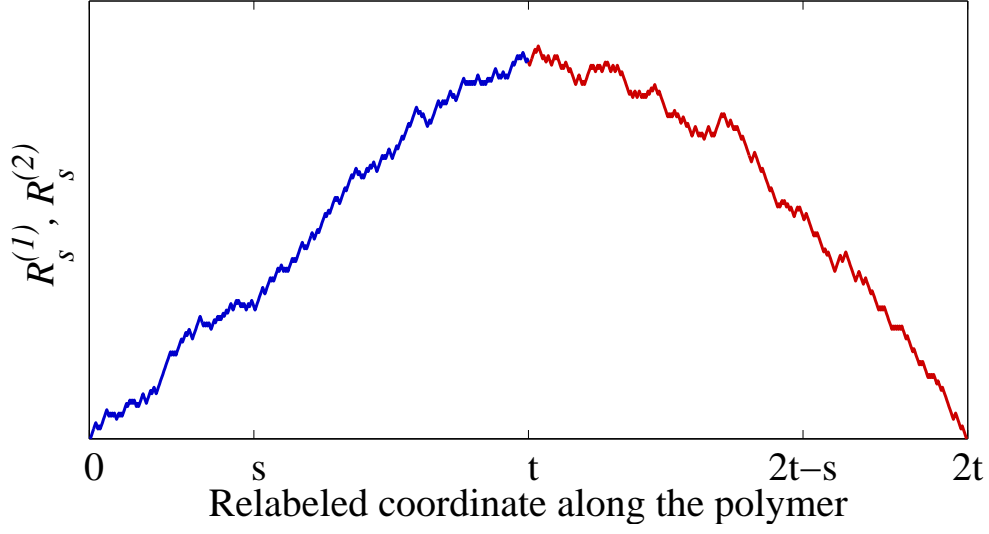Note that we subtracted the mean $m(t)$, so the real position of the unconditional

16

Figure 2: Effective sample path from $0 \to s \to t \to 2t - s \to 0$.

process is

$$R_{\text{eff};s} := \begin{cases} m(s) + \tilde{R}_{\text{eff};s}, & \text{if } s \leq t, \\ m(2t - s) + \tilde{R}_{\text{eff};s}, & \text{if } s > t. \end{cases} \tag{40}$$

The zero-mean conditional effective process is then prescribed by (7)

$$\tilde{R}_{\text{eff};s}^{cond} = \tilde{R}_{\text{eff};s} - \frac{\sigma^2_{\text{eff};0 \to s/a}}{\sigma^2_{\text{eff};0 \to L'/a}} \tilde{R}_{\text{eff};L'} \tag{41}$$

and in real space the conditional process is

$$R_{\text{eff};s}^{cond} = \begin{cases} m(s) + \tilde{R}_{\text{eff};s}^{cond}, & \text{if } s \leq t, \\ m(2t - s) + \tilde{R}_{\text{eff};s}^{cond}, & \text{if } s > t. \end{cases} \tag{42}$$

Now, denote the variances between points of interest $0 \leq s \leq t \leq 2t - s \leq L'$

17

to be

$$a_1 := \sigma^2_{\text{eff};0 \to s/a} = \sigma^2_{\text{eff};(L'-s)/a \to L'}, \tag{43a}$$

$$a_2 := \sigma^2_{\text{eff};s/a \to t/a} = \sigma^2_{\text{eff};t \to (L'-s)/a}, \tag{43b}$$

we can easily compute the variance of $R^{cond}_{\text{eff};s} - R^{cond}_{\text{eff};2t-s}$ for any $s < t$

$$\text{var}\left[R^{cond}_{\text{eff};s} - R^{cond}_{\text{eff};2t-s}\right]$$

$$= \mathbb{E}\left[\left(\tilde{R}_{\text{eff};s} - \tilde{R}_{\text{eff};2t-s} - \frac{\sigma^2_{\text{eff};0 \to s/a} - \sigma^2_{\text{eff};0 \to (L'-s)/a}}{\sigma^2_{\text{eff};0 \to L'}} \tilde{R}_{\text{eff};L'}\right)^2\right] \tag{44a}$$

$$= \mathbb{E}\left[\left(\tilde{R}_{\text{eff};s} - \tilde{R}_{\text{eff};2t-s} - \frac{\sigma^2_{\text{eff};s/a \to (L'-s)/a}}{\sigma^2_{\text{eff};0 \to L'}} \tilde{R}_{\text{eff};L'}\right)^2\right] \tag{44b}$$

$$= \mathbb{E}\left[\left(\tilde{R}_{\text{eff};s} - \tilde{R}_{\text{eff};2t-s} - \gamma \tilde{R}_{\text{eff};L'}\right)^2\right] \tag{44c}$$

where $\gamma := 2a_2/(a_1 + a_2)$. By utilizing independent increments of the processe,

$$\mathbb{E}\left[\tilde{R}_{\text{eff};s_1} \tilde{R}_{\text{eff};s_2}\right] \equiv \sigma^2_{\text{eff};0 \to \min\{s_1,s_2\}}, \tag{45}$$

we can expand (44c) and it is very easy to compute the variance of the distance

betweeen homologous sites s and $L' - s$ ($0 \le s \le t$):

$$\text{var}\left[R^{cond}_{\text{eff};s} - R^{cond}_{\text{eff};L'-s}\right]$$

$$= \mathbb{E}\left\{\tilde{R}^2_{\text{eff};s} + \tilde{R}^2_{\text{eff};L'-s} + \gamma^2 \tilde{R}^2_{\text{eff};L'}\right.$$

$$\left. + 2\left(-\tilde{R}_{\text{eff};s}\tilde{R}_{\text{eff};L'-s} - \gamma\tilde{R}_{\text{eff};s}\tilde{R}_{\text{eff};L'} + \gamma\tilde{R}_{\text{eff};L'-s}\tilde{R}_{\text{eff};L'}\right)\right\} \tag{46a}$$

$$= \left\{(a_1) + (a_1 + 2a_2) + \gamma^2(a_1 + a_2) - 2\left[(a_1) - \gamma(a_1) + \gamma(a_1 + 2a_2)\right]\right\} \tag{46b}$$

$$= 2\frac{a_1 a_2}{a_1 + a_2} = 2\frac{\sigma^2_{0\to s/a}\sigma^2_{0\to t/a}}{\sigma^2_{0\to t/a}}. \tag{46c}$$

The calculation demonstrate a much more concise machinery to obtain convergent results from the straightforward computation from building the joint probability distribution (aside from the effort to formulate the effective process). We finally remark that a similar calculation can be adopted to compute the covariance of the processes, which is another important measure of the Gaussian process.