Supplementary information: Circuit topology of linear polymers: a statistical mechanical treatment

Alireza Mashaghi[a], Abolfazl Ramezanpour[b]

[a] *Harvard Medical School, Harvard University, Boston, Massachusetts, USA and*

[b] *Department of Physics, University of Neyshabur, Neyshabur, Iran**

* Email: Alireza_Mashaghi-Tabari@meei.harvard.edu

Appendix A: Details of the belief propagation equations

We start from the partition function

$$Z(\lambda_p,\lambda_s,\lambda_x)=\frac{1}{M!}\sum_{\mathbf{e}}\prod_{l<l'}\left(\delta_{e_l\neq e_{l'}}e^{\tilde{\lambda}_p\delta_{\mathsf{q}(e_l;e_{l'}),p}+\tilde{\lambda}_s\delta_{\mathsf{q}(e_l;e_{l'}),s}+\tilde{\lambda}_x\delta_{\mathsf{q}(e_l;e_{l'}),x}}\right), \qquad \text{(A1)}$$

which is a weighted sum over the link configurations $\mathbf{e}$ satisfying the perfect matching constraints. We divided the right hand side by $M!$ to cancel the overcounting resulted from different link permutations. For large M, the partition function can be rewritten as

$$\begin{aligned} Z(\lambda_p,\lambda_s,\lambda_x) &= e^{(M\ln M)\phi(\lambda_p,\lambda_s,\lambda_x)} \\ &\approx \int dn_p dn_s dn_x e^{(M\ln M)[S(n_p,n_s,n_x)+\lambda_p n_p+\lambda_s n_s+\lambda_x n_x]}. \end{aligned} \qquad \text{(A2)}$$

Here $e^{(M\ln M)S(n_p,n_s,n_x)}$ is the number of matchings of given densities $n_{p,s,x}=N_{p,s,x}/N$. Note that the total number of perfect matchings is $e^{\ln(2M)!-\ln M!-M\ln 2}$ which for large M scales as $e^{M\ln M-M(1-\ln 2)}$. Moreover, we have $\tilde{\lambda}_{p,s,x}=2\frac{\ln M}{M-1}\lambda_{p,s,x}$.

We will use the Bethe approximation to compute $\phi(\lambda_p,\lambda_s,\lambda_x)$, and then by the Legendre transformation we will obtain the entropy function,

$$S(n_p^*,n_s^*,n_x^*)=\phi(\lambda_p,\lambda_s,\lambda_x)-\lambda_p n_p^*-\lambda_s n_s^*-\lambda_x n_x^*. \qquad \text{(A3)}$$

The values $n_{p,s,x}^*$ are determined by the saddle-point equations,

$$n_p^*=\langle n_p\rangle=\frac{\partial\phi}{\partial\lambda_p}=\frac{2}{M(M-1)}\sum_{l<l'}\langle\delta_{\mathsf{q}(e_l,e_{l'}),p}\rangle, \qquad \text{(A4)}$$

and similarly for $n_{s,x}^*$.

The central quantities in the Bethe approximation are the cavity marginals $\mu_{l\to l'}(e_l)$, giving the probability of having endpoints e_l for link l in the absence of interactions and constraints involving $e_{l'}$ [1, 2]. The cavity marginal $\mu_{l\to l'}(e_l)$ is obtained by considering the cavity messages from the other variables $\mu_{l''\to l}(e_{l''})$, and the local constraints depending on the $(e_l,e_{l''})$,

$$\mu_{l\to l'}(e_l)\propto\prod_{l''\neq l,l'}\left(\sum_{e_{l''}\neq e_l}e^{\tilde{\lambda}_p\delta_{\mathsf{q}(e_l;e_{l''}),p}+\tilde{\lambda}_s\delta_{\mathsf{q}(e_l;e_{l''}),s}+\tilde{\lambda}_x\delta_{\mathsf{q}(e_l;e_{l''}),x}}\mu_{l''\to l}(e_{l''})\right). \qquad \text{(A5)}$$

We solve the equations by iteration, starting from random initial marginals and updating the $\mu_{l\to l'}(e_l)$ in a random sequential way according to the above equations.

Having the cavity marginals, the two-link marginals read

$$\mu_{l,l'}(e_l, e_{l'}) \propto \delta_{e_{l'} \neq e_l} e^{\tilde{\lambda}_p \delta_{\mathsf{q}(e_l;e_{l'}),p} + \tilde{\lambda}_s \delta_{\mathsf{q}(e_l;e_{l'}),s} + \tilde{\lambda}_x \delta_{\mathsf{q}(e_l;e_{l'}),x}} \mu_{l \to l'}(e_l) \mu_{l' \to l}(e_{l'}). \tag{A6}$$

The free energy in the Bethe approximation is given by $(M \ln M)\phi = \sum_l \Delta\phi_l - \sum_{l<l'} \Delta\phi_{ll'} - \ln M!$, where $\Delta\phi_l$ and $\Delta\phi_{ll'}$ are the local free energy shifts [1]. These are the changes in the free energy after adding the constraints and the energy terms depending on e_l, and those that involve $(e_l, e_{l'})$, namely,

$$e^{\Delta\phi_l} = \sum_{e_l} \prod_{l' \neq l} \left(\sum_{e_{l'} \neq e_l} e^{\tilde{\lambda}_p \delta_{\mathsf{q}(e_l;e_{l'}),p} + \tilde{\lambda}_s \delta_{\mathsf{q}(e_l;e_{l'}),s} + \tilde{\lambda}_x \delta_{\mathsf{q}(e_l;e_{l'}),x}} \mu_{l' \to l}(e_{l'}) \right), \tag{A7}$$

$$e^{\Delta\phi_{ll'}} = \sum_{e_l \neq e_{l'}} e^{\tilde{\lambda}_p \delta_{\mathsf{q}(e_l;e_{l'}),p} + \tilde{\lambda}_s \delta_{\mathsf{q}(e_l;e_{l'}),s} + \tilde{\lambda}_x \delta_{\mathsf{q}(e_l;e_{l'}),x}} \mu_{l \to l'}(e_l) \mu_{l' \to l}(e_{l'}). \tag{A8}$$

Here the links are equivalent, so we rewrite the BP equations as

$$\mu(e_l) \propto \left(\sum_{e_{l'} \neq e_l} e^{\tilde{\lambda}_p \delta_{\mathsf{q}(e_l;e_{l'}),p} + \tilde{\lambda}_s \delta_{\mathsf{q}(e_l;e_{l'}),s} + \tilde{\lambda}_x \delta_{\mathsf{q}(e_l;e_{l'}),x}} \mu(e_{l'}) \right)^{M-2}. \tag{A9}$$

Let us represent e_l by its first endpoint and its length (e, r), then the above equation reads

$$\mu(e, r) \propto [e^{\tilde{\lambda}_p} w_p(e, r) + e^{\tilde{\lambda}_s} w_s(e, r) + e^{\tilde{\lambda}_x} w_x(e, r)]^{M-2}. \tag{A10}$$

where

$$w_p(e, r) = \sum_{e'=1}^{e-1} \sum_{r'=e+r+1-e'}^{2M-e'} \mu(e', r') + \sum_{e'=e+1}^{e+r-2} \sum_{r'=1}^{e+r-1-e'} \mu(e', r'), \tag{A11}$$

$$w_s(e, r) = \sum_{e'=1}^{e-2} \sum_{r'=1}^{e-e'-1} \mu(e', r') + \sum_{e'=e+r+1}^{2M-1} \sum_{r'=1}^{2M-e'} \mu(e', r'), \tag{A12}$$

$$w_x(e, r) = \sum_{e'=1}^{e-1} \sum_{r'=e-e'+1}^{e+r-1-e'} \mu(e', r') + \sum_{e'=e+1}^{e+r-1} \sum_{r'=e+r+1-e'}^{2M-e'} \mu(e', r'). \tag{A13}$$

Similarly, we obtain

$$e^{\Delta\phi_l} = \sum_{e=1}^{2M-1} \sum_{r=1}^{2M-e} [e^{\tilde{\lambda}_p} w_p(e, r) + e^{\tilde{\lambda}_s} w_s(e, r) + e^{\tilde{\lambda}_x} w_x(e, r)]^{M-1}, \tag{A14}$$

$$e^{\Delta\phi_{ll'}} = \sum_{e=1}^{2M-1} \sum_{r=1}^{2M-e} \mu(e, r)[e^{\tilde{\lambda}_p} w_p(e, r) + e^{\tilde{\lambda}_s} w_s(e, r) + e^{\tilde{\lambda}_x} w_x(e, r)]. \tag{A15}$$

Moreover, we have

$$\langle n_p \rangle = e^{-\Delta\phi_{ll'}} e^{\tilde{\lambda}_p} \sum_{e=1}^{2M-1} \sum_{r=1}^{2M-e} \mu(e, r) w_p(e, r). \tag{A16}$$

1. An alternative representation

We may as well use the matching variables $c_{ij} \in \{0,1\}$, showing the connectivity of nodes i and j, to rewrite the partition function A1 as

$$Z(\lambda_p, \lambda_s, \lambda_x) = \sum_{\mathbf{c}} \prod_i \delta_{\sum_{j\neq i} c_{ij}=1} \prod_{(ij)<(kl)} e^{c_{ij}c_{kl}[\tilde{\lambda}_p \delta_{\mathsf{q}(ij;kl),p} + \tilde{\lambda}_s \delta_{\mathsf{q}(ij;kl),s} + \tilde{\lambda}_x \delta_{\mathsf{q}(ij;kl),x}]}, \tag{A17}$$

This representation of the problem is more efficient than the one we used above but the BP equations are more involved; we have to distinguish between two kinds of BP message $\mu_{(ij)\to i}(c_{ij})$ and $\mu_{(ij)\to(kl)}(c_{ij})$. The former is the probability of c_{ij} in absence of the matching constraint $I_i(c_{\partial i}) \equiv \delta_{\sum_{j\neq i} c_{ij}=1}$, and the latter is computed in absence of two-link interaction $w_{(ij),(kl)}(c_{ij}, c_{kl}) \equiv \exp(c_{ij}c_{kl}[\tilde{\lambda}_p \delta_{\mathsf{q}(ij;kl),p} + \tilde{\lambda}_s \delta_{\mathsf{q}(ij;kl),s} + \tilde{\lambda}_x \delta_{\mathsf{q}(ij;kl),x}])$. Here $c_{\partial i} \equiv \{c_{ij} : j \neq i\}$. The BP equations governing these cavity marginals are

$$\mu_{(ij)\to j}(c_{ij}) \propto \left(\sum_{c_{\partial i\setminus j}} I_i(c_{\partial i}) \prod_{k\in\partial i\setminus j} \mu_{(ik)\to i}(c_{ik}) \right) \times \prod_{(kl):k,l\neq i,j} \left(\sum_{c_{kl}} w_{(ij),(kl)}(c_{ij}, c_{kl}) \mu_{(kl)\to(ij)}(c_{kl}) \right), \tag{A18}$$

and

$$\mu_{(ij)\to(kl)}(c_{ij}) \propto \left(\sum_{c_{\partial i\setminus j}} I_i(c_{\partial i}) \prod_{k\in\partial i\setminus j} \mu_{(ik)\to i}(c_{ik}) \right) \left(\sum_{c_{\partial j\setminus i}} I_j(c_{\partial j}) \prod_{k\in\partial j\setminus i} \mu_{(jk)\to j}(c_{jk}) \right) \times \prod_{(k'l')\neq(kl):k',l'\neq i,j} \left(\sum_{c_{k'l'}} w_{(ij),(k'l')}(c_{ij}, c_{k'l'}) \mu_{(k'l')\to(ij)}(c_{k'l'}) \right). \tag{A19}$$

Similarly, we can compute the one-link marginals $\mu_{(ij)}(c_{ij})$ and the two-link marginals $\mu_{(ij),(kl)}(c_{ij}, c_{kl})$.

Appendix B: Details of the minsum equations

Consider a system of interacting site variables $\sigma_i \in \{0,1\}$ for $i = 1,\dots,L$, with energy function $\mathcal{E}(\boldsymbol{\sigma}) = \sum_{i<j} \mathcal{E}_{ij}(\sigma_i, \sigma_j)$, where

$$\mathcal{E}_{ij}(\sigma_i, \sigma_j) = -[\sigma_i\sigma_j + (1-\sigma_i)(1-\sigma_j)] \ln \alpha_{ij} - [1 - \sigma_i\sigma_j - (1-\sigma_i)(1-\sigma_j)] \ln(1-\alpha_{ij}). \tag{B1}$$

The α_{ij} are here parameters, giving the probability of having a link connecting sites i, j.

We start from the finite-temperature BP equations for the cavity marginals of the probability distribution of variable configurations $\mathcal{P}(\boldsymbol{\sigma}) \propto e^{-\beta\mathcal{E}(\boldsymbol{\sigma})}$,

$$\nu_{i\to j}(\sigma_i) \propto \prod_{k\neq i,j}\left(\sum_{\sigma_k} e^{-\beta\mathcal{E}_{ik}(\sigma_i,\sigma_k)}\nu_{k\to i}(\sigma_k)\right). \tag{B2}$$

This is the probability of state σ_i for site i in the absence of interaction with site j. It is more appropriate to work with the cavity fields $h_{i\to j} \equiv \frac{1}{\beta}\ln\left(\frac{\nu_{i\to j}(1)}{\nu_{i\to j}(0)}\right)$ where the BP equations read

$$\beta h_{i\to j} = \sum_{k\neq i,j}\ln\left(e^{-\beta\mathcal{E}_{ik}(1,0)} + e^{-\beta\mathcal{E}_{ik}(1,1)+\beta h_{k\to i}(\sigma_k)}\right) - \sum_{k\neq i,j}\ln\left(e^{-\beta\mathcal{E}_{ik}(0,0)} + e^{-\beta\mathcal{E}_{ik}(0,1)+\beta h_{k\to i}(\sigma_k)}\right). \tag{B3}$$

Now take the limit $\beta \to \infty$ of the above equations. The resulting equations for the cavity messages $h_{i\to j}$ are called minsum equations [2] and read

$$h_{i\to j} = \sum_{k\neq i,j}\max(\ln(1-\alpha_{ik}), \ln\alpha_{ik} + h_{k\to i}) - \sum_{k\neq i,j}\max(\ln\alpha_{ik}, \ln(1-\alpha_{ik}) + h_{k\to i}). \tag{B4}$$

We solve the minsum equations for the cavity messages by iteration, starting from random initial messages. In the end, the local messages h_i are obtained like the cavity ones but considering all the incoming messages from the neighboring variables.

To find a configuration minimizing the energy, we use the reinforcement algorithm [3]: In each step of updating the cavity messages, we add external fields that polarize the messages in the direction suggested by the local messages. More precisely, the reinforced minsum equations read

$$h_{i\to j}^{t+1} = r(t)h_i^t + \sum_{k\neq i,j}\max(\ln(1-\alpha_{ik}), \ln\alpha_{ik} + h_{k\to i}^t) - \sum_{k\neq i,j}\max(\ln\alpha_{ik}, \ln(1-\alpha_{ik}) + h_{k\to i}^t). \tag{B5}$$

Similarly, we update the local messages

$$h_i^{t+1} = r(t)h_i^t + \sum_{k\neq i}\max(\ln(1-\alpha_{ik}), \ln\alpha_{ik} + h_{k\to i}^t) - \sum_{k\neq i}\max(\ln\alpha_{ik}, \ln(1-\alpha_{ik}) + h_{k\to i}^t). \tag{B6}$$

The reinforcement parameter $r(t)$ is zero at the beginning of the algorithm ($t = 0$) and grows slowly by t, for example as $r(t+1) = r(t) + \delta$.

Appendix C: More details and figures

Here we present more details of the numerical simulations and figures obtained in this study.

Figure 1 displays the approximate entropy (logarithm of the number of link configurations $\mathcal{N}$) that is obtained within the Bethe approximation. Here we take $\lambda_x = 0$ and report the entropy in the space of parameters (λ_p, λ_s) even if the algorithm does not converges. The solution to the BP equations A9 is found by iteration and the algorithm converges when the difference in the BP messages $\mu(e_l)$ in two successive steps of the iteration is less than a convergence limit $\epsilon = 10^{-8}$. In the same figure, we observe the region in the parameter space that the BP algorithm converges. Given the BP messages, the entropy is computed by Eqs. A3 and A7,A8.

Figures 2, 3, and 4 show the one-link and two-link distributions for more parameter samples obtained by the BP algorithm as described above. In Fig. 5, we compare the two-link distance distribution obtained by the approximate algorithm with the exact one for a small number of links.

In Fig. 6, we compare the reconstructed one-link and two-link distributions with the observed data from link configurations with a regular sector of size $L/2$. The inferred statistics can be improved by iteration using the information obtained in the previous stages of the algorithm. In the figure, we also compare the model data obtained without any prior information (a), and with the information provided in the first stage of the algorithm (b).

---

[1] M. Mézard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, Oxford, 2009).

[2] F. R. Kschischang, B. J. Frey, and H. -A. Loeliger, IEEE Trans. Infor. Theory **47**, 498 (2001).

[3] A. Braunstein and R. Zecchina, Phys. Rev. Lett. **96**, 030201 (2006).

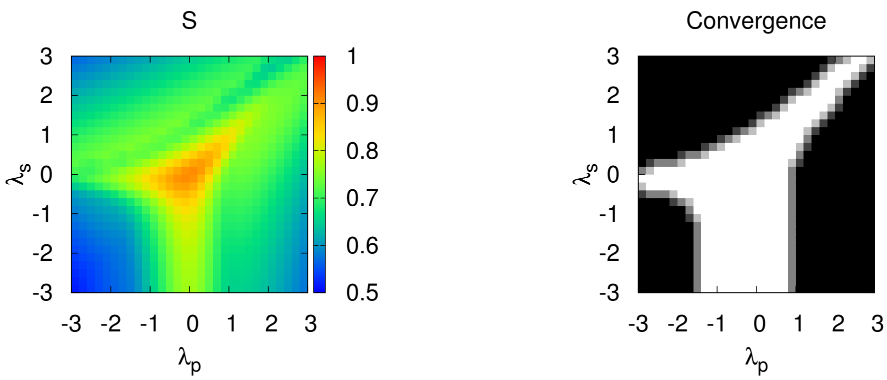

FIG. 1. The entropy $S = \frac{1}{M \ln M} \ln \mathcal{N}$ ($\mathcal{N}$ is the number of configurations) obtained by the Bethe approximation, and the region that the BP algorithm converges (white region). The data are for $M = 40$ links and $\lambda_x = 0$. The convergence limit is $\epsilon = 10^{-8}$.

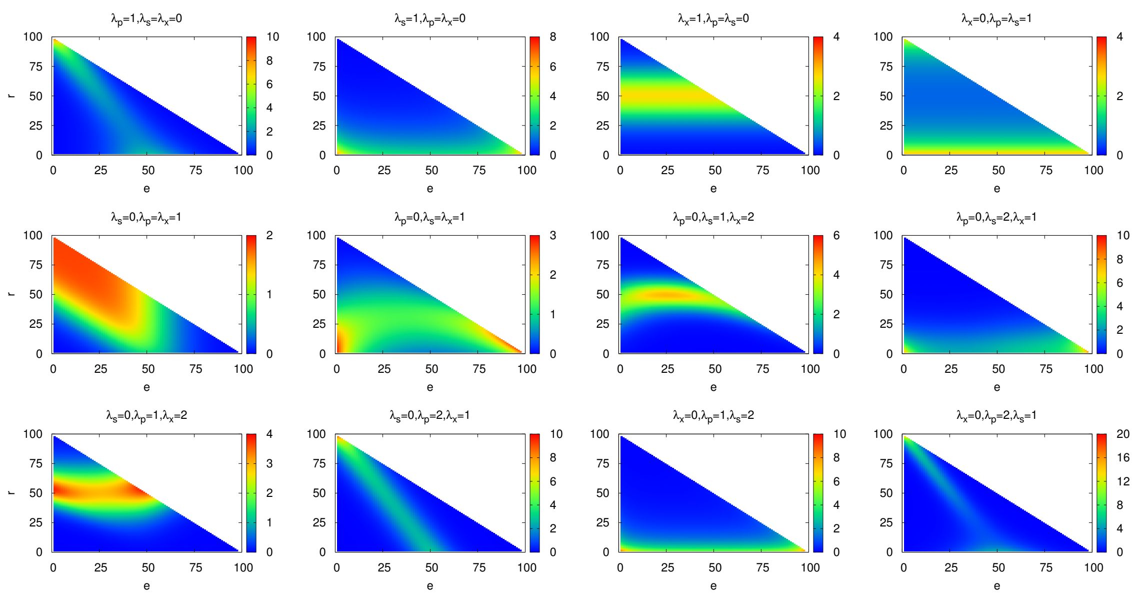

FIG. 2. One-link distribution (more precisely $M(2M-1)\mu(e,r)$) obtained by the Bethe approximation for different energy parameters $\lambda_{p,s,x}$ with $M = 50$ links. Here $\mu(e,r)$ is the probability of having a link with the first endpoint e and length r.

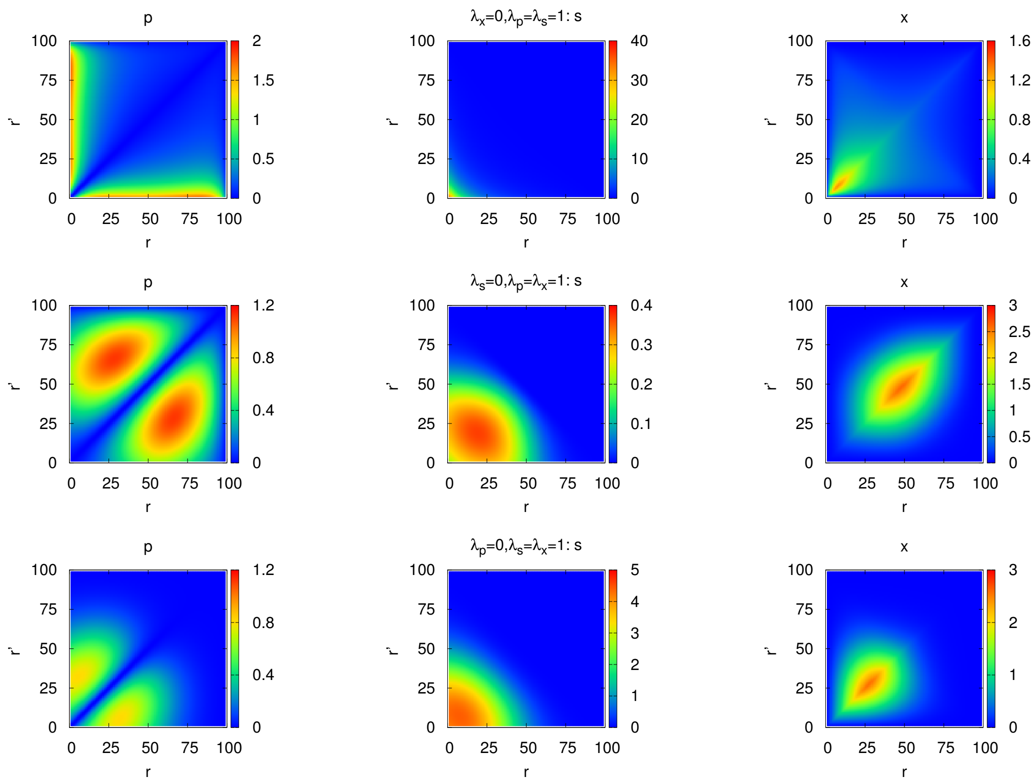

FIG. 3. Two-link length distribution $\mu_{ll'}(r,r')$ (multiplied by a constant to make it of order one) obtained by the Bethe approximation for $M = 50$ links. Here $\mu_{ll',q}(r,r')$ is the probability of finding two-links of type $q = p, s, x$ with lengths r and r'. The energy parameters $\lambda_{p,s,x}$ and the average two-link densities are fixed in each row. The columns are for different types of two-links: p (left), s (center), and x (right).

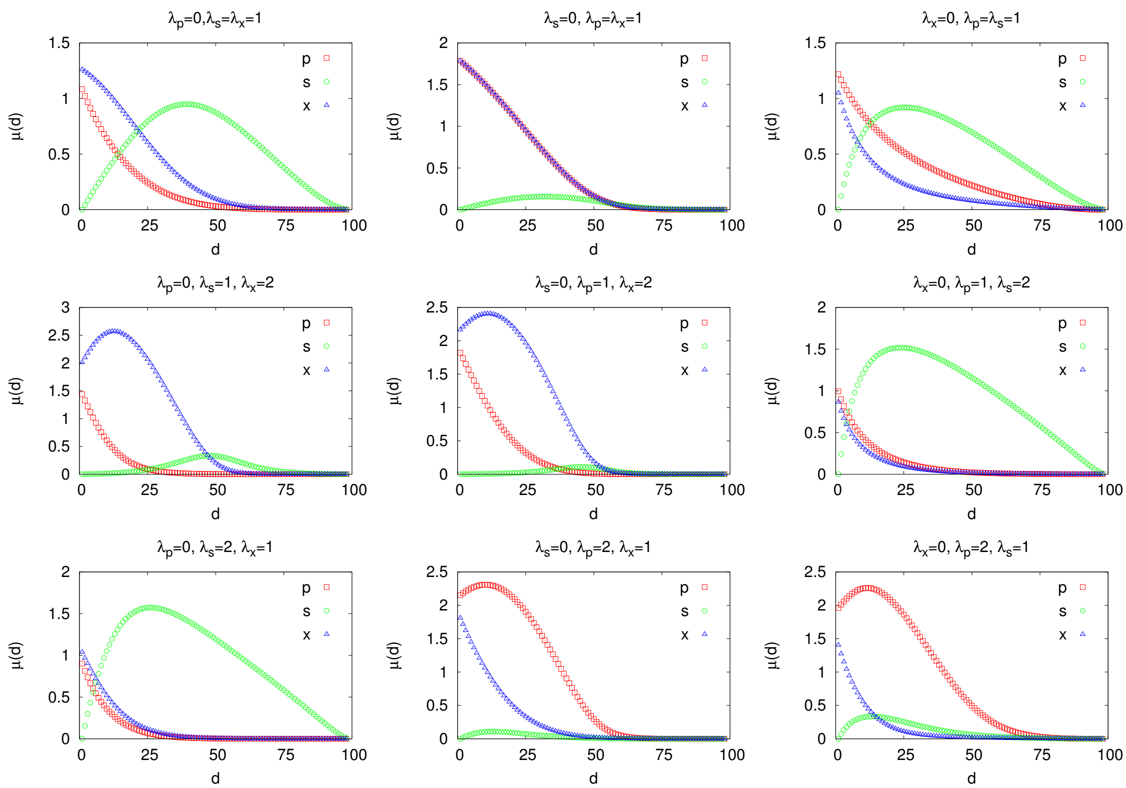

FIG. 4. Two-link distance distribution $\mu_{ll'}(d)$ (multiplied by a constant to make it of order one) for different energy parameters $\lambda_{p,s,x}$ and different types of two-links (p, s, x), obtained by the Bethe approximation for $M = 50$ links. Here $\mu_{ll',q}(d)$ is the probability of finding two links of type $q = p, s, x$ at distance d (separation of the first endpoints) from each other.

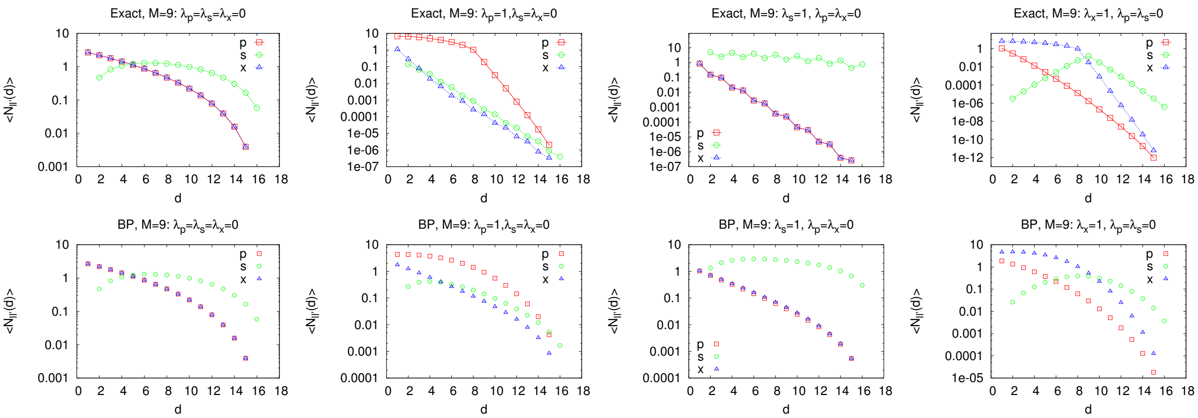

FIG. 5. Comparing the average two-link numbers $\langle N_{ll'}(d)\rangle$ obtained exactly (top) with those of the Bethe approximation (bottom) for $M = 9$ links. Distance d of two links is the separation of their first endpoints. Each panel shows $\langle N_{ll'}(d)\rangle$ for fixed energy parameters $\lambda_{p,s,x}$ but different types p, s, x.

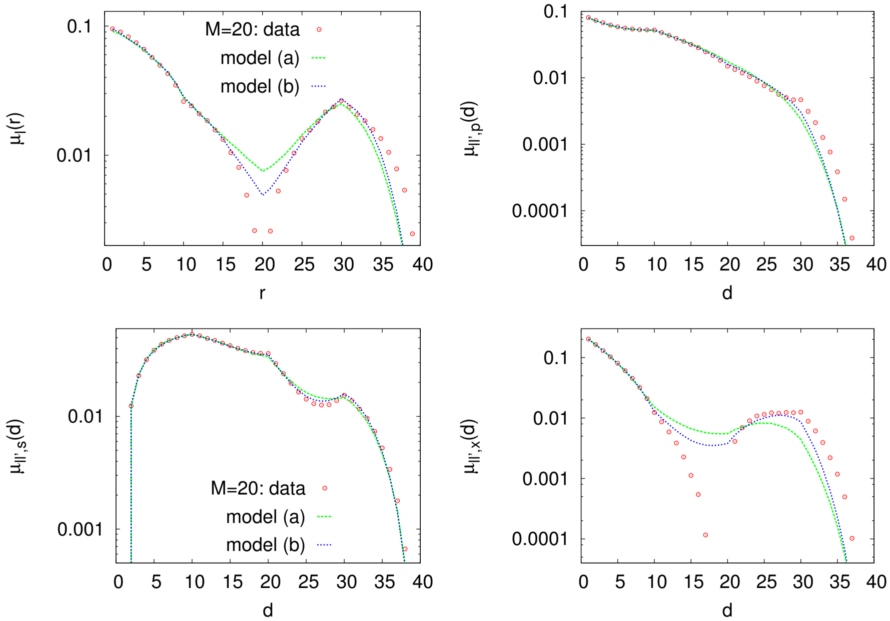

FIG. 6. The one- and two-link probability distributions $\mu_l(r), \mu_{ll',p,s,x}(d)$ obtained by the inverse algorithm given the average numbers $M^*(r), N^*_{p,s,x}(d)$ extracted from 10000 randomly generated configurations of $M = 20$ links in presence of a regular sector of size $L/2$ in the center of the chain. Hear r refers to the length of link, and d gives the distance between the first endpoints of two links. $M^*(r)$ and $N^*_{p,s,x}(d)$ are the average number of links of length r and two-links of distance d, respectively. Model (a) is obtained by running the inverse algorithm with no prior information of the sector. To obtain model (b), we run the inverse algorithm with an additional external field disfavoring some connections according to the one-link probability distribution $\mu_l(e, r)$ provided by model (a). $\mu_l(e, r)$ is the probability of having a link with first endpoint e and length r.