Freezing transition of the random bond RNA model: Statistical properties of the pairing weights

Cécile Monthus and Thomas Garel
Service de Physique Théorique, CEA/DSM/SPhT, Unité de recherche associée au CNRS 91191 Gif-sur-Yvette cedex, France
(Received 27 November 2006; published 12 March 2007)

To characterize the pairing specificity of RNA secondary structures as a function of temperature, we analyze the statistics of the pairing weights as follows: for each base (i) of the sequence of length N, we consider the $(N-1)$ pairing weights $w_i(j)$ with the other bases $(j \neq i)$ of the sequence. We numerically compute the probability distributions $P_1(w)$ of the maximal weight $w_i^{\max}=\max_j[w_i(j)]$, the probability distribution $\Pi(Y_2)$ of the parameter $Y_2(i)=\Sigma_j w_i^2(j)$, as well as the average values of the moments $Y_k(i)=\Sigma_j w_i^k(j)$. We find that there are two important temperatures $T_c<T_{\rm gap}$. For $T>T_{\rm gap}$, the distribution $P_1(w)$ vanishes at some value $w_0(T)<1$, and accordingly the moments $\overline{Y_k(i)}$ decay exponentially as $[w_0(T)]^k$ in k. For $T<T_{\rm gap}$, the distributions $P_1(w)$ and $\Pi(Y_2)$ present the characteristic Derrida-Flyvbjerg singularities at $w=1/n$ and $Y_2=1/n$ for $n=1,2,\ldots$. In particular, there exists a temperature-dependent exponent $\mu(T)$ that governs the singularities $P_1(w)\sim(1-w)^{\mu(T)-1}$ and $\Pi(Y_2)\sim(1-Y_2)^{\mu(T)-1}$ as well as the power-law decay of the moments $\overline{Y_k(i)}\sim 1/k^{\mu(T)}$. The exponent $\mu(T)$ grows from the value $\mu(T=0)=0$ up to $\mu(T_{\rm gap})\sim 2$. The study of spatial properties indicates that the critical temperature T_c where the large-scale roughness exponent changes from the low temperature value $\zeta\sim 0.67$ to the high temperature value $\zeta\sim 0.5$ corresponds to the exponent $\mu(T_c)=1$. For $T<T_c$, there exists frozen pairs of all sizes, whereas for $T_c<T<T_{\rm gap}$, there exists frozen pairs, but only up to some characteristic length diverging as $\xi(T)\sim 1/(T_c-T)^{\nu}$ with $\nu\simeq 2$. The similarities and differences with the weight statistics in Lévy sums and in Derrida's random energy model are discussed.

DOI: 10.1103/PhysRevE.75.031103 PACS number(s): 02.50.−r, 87.14.Gg, 87.15.Cc, 64.70.Pf

I. INTRODUCTION

Models of RNA secondary structures [1,2] have been recently studied by physicists [3–11], because of the similarity of the low-temperature disorder-dominated phase with the spin glass phase. This phase has been analyzed either in terms of replica theory of spin glasses [12], with nontrivial overlap distribution function $P(q)$ [3,7,9], or in terms of the droplet theory of spin glasses [13]: in this case, a finite droplet exponent $\theta>0$ has been obtained via the ϵ-coupling method, with finite values $\theta\sim 0.23$ [5,9], $\theta=1/3$ [8] whereas a vanishing droplet exponent $\theta=0$ and logarithmic droplets have been found via the statistics of pinch free energies [6,11]. Many authors have also studied the phase transition towards the high-temperature molten phase, with different values for the correlation length exponent ν and the specific heat exponent $\alpha=2-\nu$. Numerical results have given, for instance, $\nu\sim 3.9$ [7] or $\nu\sim 1.1$ [9], and the field theory of [10] predicts $\nu\sim 8/5\sim 1.6$, whereas the general theorem of Ref. [14] on phase transitions in disordered systems states that the finite-size correlation exponent ν has to satisfy the bound $\nu\geqslant 2/d=2$.

In this paper, we try to clarify the nature of the low temperature phase and of the freezing transition by studying statistical properties of the pairing weights. The paper is organized as follows. The model and usual observables are recalled in Sec. II. We then present a detailed study of the pairing weights seen by a given monomer. For clarity, the statistical properties of the weights alone, independently of the distances involved are described in Sec. III, whereas the study of spatial properties is given in Sec. IV. We summarize our results in Sec. V. For comparison, we recall in the Appendix the properties of the weights statistics in Lévy sums and in Derrida's random energy model (REM), as well as the corresponding Derrida-Flyvberg singularities.

II. MODEL AND OBSERVABLES

A. Partition function

An RNA secondary structure of a sequence of N bases $(1,2,\ldots,N)$ is a set of base pairs all compatible with each other. To be compatible, two pairs (i,j) and (k,l) have to be nonoverlapping (for instance, $i<j<k<l$) or nested (for instance, $i<k<l<j$) [1]. The energy of an allowed configuration $\mathcal{C}$ is then the sum of the energies $\epsilon_{i,j}$ of all the pairs (i,j) that are present in the configuration

$$E(\mathcal{C})=\sum_{(i,j)\in\mathcal{C}}\epsilon_{i,j}. \tag{1}$$

This noncrossing property of pairs allows us to write the following recursion for the partition functions $Z_{i,j}$ of intervals $(i,i+1,\ldots j-1,j)$ [1]:

$$Z_{i,j}=Z_{i,j-1}+\sum_{k=i}^{j-p_{\min}} Z_{i,k-1}e^{-\beta\epsilon_{k,j}}Z_{k+1,j-1}. \tag{2}$$

The first terms represent the configurations where j is unpaired, whereas the second term represents the configurations where j is paired with the base $k\in\{i,i+1,\ldots,j-p_{\min}\}$, and $p_{\min}$ represents the minimal distance along the sequence to form a pair. So the full partition function $Z_{1,N}$ can be computed in a CPU time of order $O(N^3)$. In the literature [3–11], various choices for the parameter $p_{\min}$ and for the distribution of the energies $\epsilon_{i,j}$ have been made. To mimic real RNA sequences, some authors have chosen to work with $p_{\min}=4$ and random sequences of the four bases A,C,G,U, where only A-U and G-C pairs are possible [3,4]. However other authors have chosen other values of $p_{\min}=1,2$, and a two-letter code [7], or to add some continuous disorder to avoid exact degeneracies [5,8]. Instead of sequence disorder, the case of bond disorder, where the $\epsilon_{i,j}$ are independent random

variables has been used in Refs. [6,10,11]. Finally in Ref. [9], more complicated models including also stacking energies have been studied.

All the numerical results we will present below corresponds to the case $p_{\min}=1$ (the convergence towards the asymptotic regime $N\to\infty$ is then much more rapid than for the "biological value" $p_{\min}=4$), and to the bond-disorder case, where the $\epsilon_{i,j}$ are independent random variable drawn with the flat distribution

$$\rho(\epsilon)=1 \text{ for } -2.5\leqslant\epsilon\leqslant-1.5. \tag{3}$$

The sequence length N and the corresponding number $ns(N)$ of independent sequences that we have studied are typically as follows:

$$N=50,100,200,400,600,800,$$

$$ns(N)=18\times10^6,2\times10^6,3\times10^5,25\times10^3,6\times10^3,2\times10^3. \tag{4}$$

B. Pair probabilities

The pairing probability of bases (i,j) in the sequence $(1,N)$ reads

$$P_{i,j}=\frac{e^{-\beta\epsilon_{i,j}}Z_{i+1,j-1}Z^{\text{ext}}_{i,j}}{Z_{1,N}}, \tag{5}$$

where $Z_{i+1,j-1}$ represents the partition function of the internal sequence $(i+1,\ldots,j-1)$ computed in Eq. (2), and where $Z^{\text{ext}}_{i,j}$ represents the partition function of the external sequence $(1,2,\ldots,i-1,j+1,\ldots,N)$, which can be computed by extending the recursion of Eq. (2) to the duplicated sequence $(1,2,\ldots,N,1,2,\ldots,N)$: $Z^{\text{ext}}_{i,j}$ is then given by $Z_{j+1,N+i-1}$ [6]. So the pair probabilities $P_{i,j}$ can also be computed in a CPU time of order $O(N^3)$.

C. Height profile

An RNA secondary structure $\mathcal{C}$ can be represented as a noncrossing arch diagram or equivalently as a "mountain profile" (see Fig. 3 of Ref. [6]), where the height h_k represents the number of pairs (i,j) such that $i<k<j$: this height starts at $h(k=1)=0$, ends at $h(k=N)=0$, remains nonnegative in between, and the difference $(h_{k+1}-h_k)$ can only take the three values $(+1,0,-1)$. Its thermal average reads in terms of the pair probabilities of Eqs. (5)

$$\langle h_k\rangle=\sum_{i<k<j}P_{i,j}. \tag{6}$$

D. Overlap

In disorder-dominated phases, such as spin-glasses or directed polymers for instance, the overlap is usually a convenient order parameter. Here, the overlap between two configurations $\mathcal{C}_1$ and $\mathcal{C}_2$ can be defined as

$$q(\mathcal{C}_1,\mathcal{C}_2)=\frac{2}{N}\sum_{i<j}1_{(i,j)\in\mathcal{C}_1}1_{(i,j)\in\mathcal{C}_2}, \tag{7}$$

where the normalization factor $N/2$ represents the number of pairs existing in the ground state. The thermal average reads

$$q_2(T)=\langle q(\mathcal{C}_1,\mathcal{C}_2)\rangle=\frac{2}{N}\sum_{i<j}P^2_{i,j} \tag{8}$$

in terms of the pair probabilities of Eq. (5). However, this overlap is not an appropriate order parameter for RNA secondary structure, because it does never vanish, even at $T=\infty$ as we now explain.

E. Limit of infinite temperature

In the limit $T=\infty$, disorder disappears, and the partition function can be exactly computed [6]

$$Z_{1,N}(T=\infty)\underset{N\to\infty}{\propto}\frac{3^N}{N^{3/2}}. \tag{9}$$

This number of possible configurations corresponds to the number of one-dimensional (1D) random walks for the height h_k, with three choices per step for the height increment $h_{k+1}-h_k=0,\pm1$, and where the factor $1/N^{3/2}$ is a first-return probability. From this interpretation of the height as a positive random walk, it is clear that the middle height $h_{N/2}$ scales as $N^{1/2}$

$$\langle h_{N/2}\rangle(T=\infty)\sim N^{1/2}. \tag{10}$$

For $1\ll l\ll N$, the pair probability of Eq. (5) behaves as

$$P_{i,i+l}(T=\infty)\propto\frac{N^{3/2}}{l^{3/2}(N-l)^{3/2}}. \tag{11}$$

However, the overlap [Eq. (8)] is finite

$$q_2(T=\infty)>0 \tag{12}$$

because small pairs have a finite weight, in particular for $l=1$, Eq. (5) yields

$$P^{(T=\infty)}_{i,i+1}\sim\frac{Z_{1,N-2}(T=\infty)}{Z_{1,N}(T=\infty)}\sim\frac{1}{3^2}. \tag{13}$$

F. Limit of zero temperature

At $T=0$, there is a numerical consensus [5,6,11] that the disorder-averaged height has a different scaling from the random walk value of Eq. (10)

$$\overline{\langle h_{N/2}\rangle}(T=0)\sim N^{\zeta}, \tag{14}$$

where the roughness exponent

$$\zeta\sim0.67 \tag{15}$$

is extremely close to the simple value $2/3$, although we are not aware of any rigorous or heuristic argument in favor of this fraction. The exponent η governing the scaling of large pairs $1\ll l\ll N$

$$\overline{P_{i,i+l}(T=0)} \sim \frac{1}{N^{\eta}}\Phi\left(\frac{l}{N}\right) \tag{16}$$

is actually directly related to the roughness exponent via the relation [5,6,10,11]

$$\eta = 2 - \zeta \tag{17}$$

as can be seen from the definition of the height [Eq. (6)]. This relation valid at any temperature corresponds to $\eta(T=\infty)=3/2$ in agreement with Eq. (11), and to

$$\eta(T=0) \sim 1.33 \tag{18}$$

as directly measured in Refs. [5,11].

G. Characterization of the transition in previous works

Previous work has shown that there exists a finite temperature transition between a high temperature or molten phase, where entropy dominates, and a low-temperature phase where disorder dominates. But very different observables have been used numerically to characterize the transition, such as the overlap probability distribution $P(q)$ [3,7,9], the ϵ-coupling method [5], and the so-called "pinch-free energy" [6,11]. In the field theory of Ref. [10], the critical exponents exactly at criticality were found to be the same as the ones in the low temperature phase, both for the height

$$\zeta(T > T_c) = \frac{1}{2},$$

$$\zeta(T \leqslant T_c) = \zeta(T=0) \tag{19}$$

and for the overlap $\overline{P^2_{i,i+l}}$ of large pairs $l \gg 1$

$$\overline{P^2_{i,i+l}} \sim (\overline{P_{i,i+l}})^2 \sim \frac{1}{l^3} \text{ for } T > T_c,$$

$$\overline{P^2_{i,i+l}} \sim \overline{P_{i,i+l}} \sim \frac{1}{l^{\eta(T=0)}} \text{ for } T \leqslant T_c. \tag{20}$$

In the following, we propose to study the freezing transition via the statistical properties of the pair weights seen by a given monomer. In Ref. [4], the integrated probability distribution of the maximal weight $p_{\max}$ seen by a given monomer has been measured to characterize the barrier statistics between degenerate ground states for discrete disorder, but the phase transition region was not studied in details from this point of view. In another context concerning disordered polymers [15], the distribution of the maximal weight $p_{\max}$ was used to analyze a phase transition, the important region being there the neighborhood of $p_{\max} \to 0$, whereas in the present study, the important region is the region $p_{\max} \to 1$.

III. STATISTICAL PROPERTIES OF THE PAIR WEIGHTS

A. Pair landscape seen by a given monomer

For each base (i) of the sequence of length N, we consider the $(N-1)$ pairing weights with the other bases $(j \neq i)$ of the sequence [Eq. (5)]

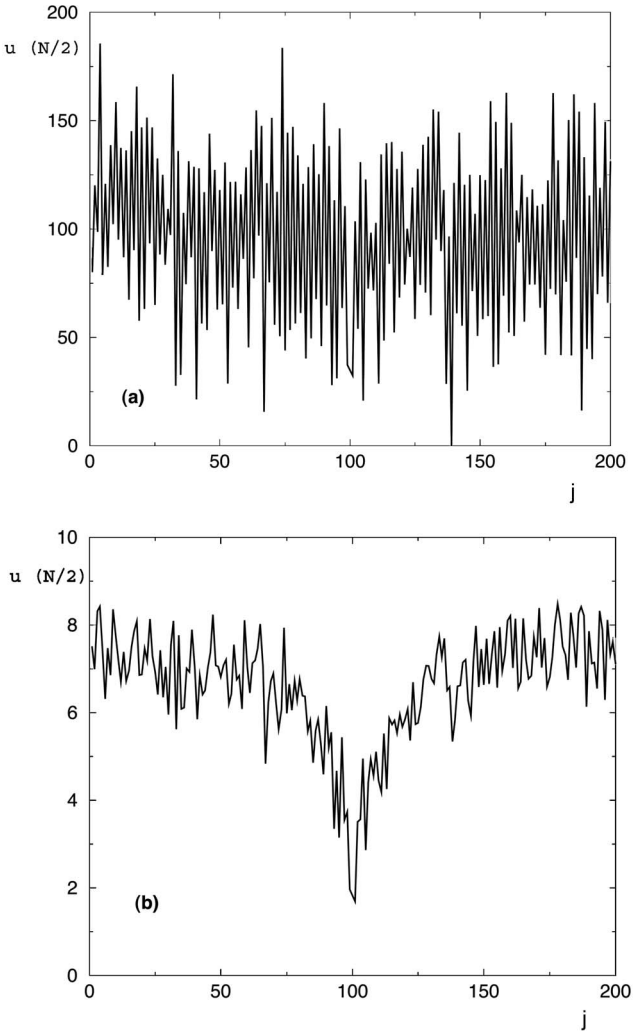


FIG. 1. Pairing weight landscape $u(\frac{N}{2})=[-\ln w_{N/2}(j)]$ seen by the middle monomer [see Eq. (21)], for $N=200$ at low and high temperatures (a) at low temperature ($T=0.02$), only a few weights dominate [note the scale of $\ln w_i(j)$], at some random positions (b) at high temperature ($T=0.5$), the disorder is a small perturbation with respect to the entropy of the pure case that favors the neighbors.

$$w_i(j) \equiv P_{i,j} = \frac{e^{-\beta\epsilon_{i,j}} Z_{i+1,j-1} Z^{\text{ext}}_{i,j}}{Z_{1,N}}. \tag{21}$$

Making the convention that $w_i(i)$ denotes the weight of the configurations where (i) is unpaired

$$w_i(i) \equiv \frac{Z^{\text{ext}}_{i,i}}{Z_{1,N}} \tag{22}$$

these weights are normalized to

$$\sum_{j \neq i} w_i(j) = 1 - w_i(i). \tag{23}$$

The pairing weight landscape seen by a given monomer is shown for two temperatures in Fig. 1: in the low-temperature frozen phase, only a few weights dominate in continuity with

the limit of zero temperature where only one weight is nonzero, whereas in the high temperature phase, many weights contribute, and disorder represents a small random correction around the entropic term of the $T=\infty$ limit [Eq. (11)].

In the rest of this section, we describe the statistical properties of the weights alone, independently of the distances involved. The study of spatial properties is postponed to Sec. IV for clarity.

B. Characterization of the weights statistics

In analogy with the weight statistics in Lévy sums and in the random energy model [16,17] (we refer the reader to the Appendix for a summary of the most important results for our present study), we have numerically computed the probability distributions $P_1(w)$ of the maximal weight

$$w_i^{\max} = \max_{j\neq i}\{w_i(j)\} \tag{24}$$

as well as $P_2(w)$ of the second maximal weight.

Another useful way to characterize the statistical properties of the weights [16,17] (see the Appendix) is to consider the moments of arbitrary order k

$$Y_k(i) = \sum_{j\neq i} w_i^k(j) \tag{25}$$

that represents the probability that the monomer (i) is paired to the same monomer in k different thermal configurations of the same disordered sample. We have measured the probability $\Pi(Y_2)$ of the parameter

$$Y_2(i) = \sum_{j\neq i} w_i^2(j) \tag{26}$$

as well as the moments $\overline{Y_k(i)}$ for $2\leqslant k\leqslant 100$. Finally, we have also computed the density of weights

$$f(w) = \overline{\sum_{j\neq i} \delta[w - w_i(j)]} \tag{27}$$

giving rise to the moments

$$\overline{Y_k(i)} = \int_0^1 dw w^k f(w). \tag{28}$$

The normalization condition for the density $f(w)$ is

$$\overline{Y_1(i)} = \int_0^1 dw w f(w) = 1. \tag{29}$$

The properties of all these quantities in the case of Lévy sums of independent variables are recalled in the Appendix. In the following, we describe their properties for the RNA case and discuss the similarities and differences with Lévy sums. The numerical results for the histograms $P_1(w)$, $P_2(w)$, $f(w)$, $\Pi(Y_2)$ have been obtained by collecting the weights seen by each monomer $i=1,2\ldots N$ in the $n_s(N)$ disordered sequences generated [see Eq. (4) for the numerical values used for N and $n_s(N)$].

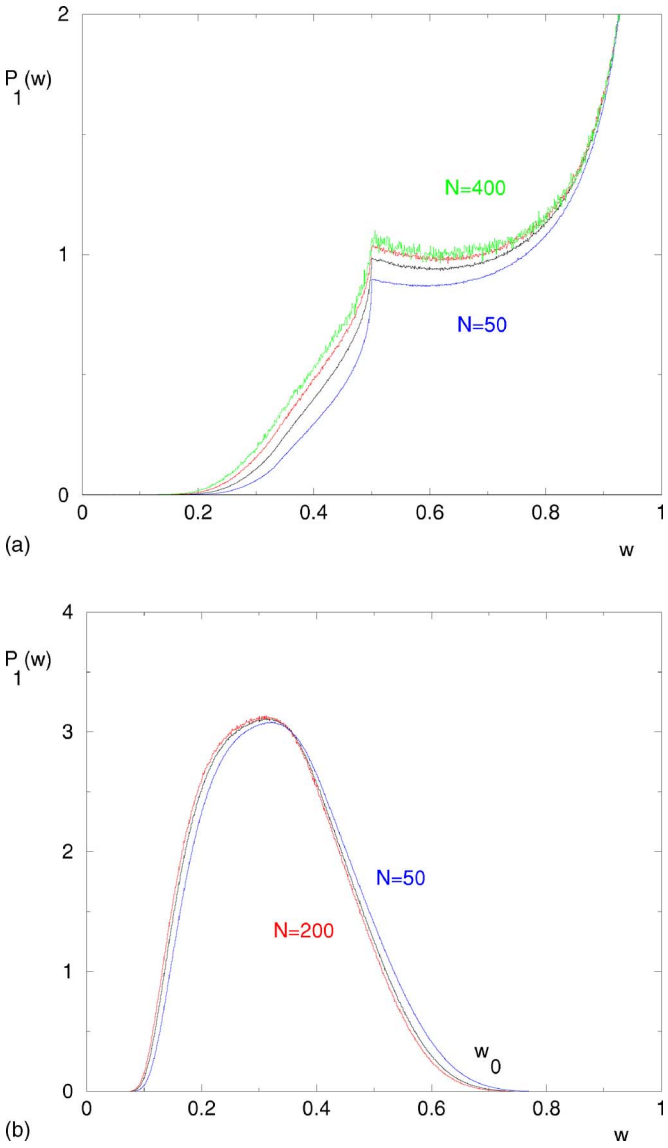


FIG. 2. (Color online) Probability distribution $P_1(w)$ of the largest weight seen by a given monomer [see Eq. (24)] (a) at $T=0.05$ (low-temperature phase) for $N=50,100,200,400$: the characteristic Derrida-Flyvbjerg singularities at $w=1$ and $w=1/2$ are clearly visible. (b) at $T=0.4$ (high-temperature phase) for $N=50,100,200$: the distribution $P_1(w)$ does not reach $w=1$ anymore, but vanishes at some maximal value $w_0(T)<1$.

C. Probability distribution $P_1(w)$ of the largest weight seen by a given monomer

The probability distribution $P_1(w)$ of the largest weight seen by a given monomer is shown in Fig. 2 for low and high temperatures. At low temperature [Fig. 2(a)], this distribution presents a divergent singularity near $w\to 1$

$$P_1(w) \underset{w\to 1}{\propto} (1-w)^{\mu(T)-1} \tag{30}$$

with a temperature dependent exponent $\mu(T)$, called μ in analogy with the case of Lévy sum of index μ, and with the random energy model where $\mu(T)=T/T_g$ (see the Appendix). However here in RNA, the pairings free-energies are not

independent variables and are not drawn with the same distribution (as a consequence of the distances involved), so the full distribution $P_1(w)$ cannot coincide with the Lévy sums result. Nevertheless, we find that $P_1(w)$ presents the characteristic Derrida-Flyvbjerg singularities at $w=1/n$ (see the Appendix). The stronger singularity occurs at $w=1$ and defines the exponent $\mu(T)$ (30), but the second singularity at $w=1/2$ is also clearly visible on Fig. 2. This shows that for each base in the frozen phase, all the weight is concentrated on a few pairing partners j.

At sufficiently high temperature [Fig. 2(b)], the distribution $P_1(w)$ does not reach $w=1$ anymore, but vanish at some maximal value $w_0(T)<1$

$$P_1(w) \underset{w\to w_0(T)}{\propto} [w_0(T)-w]^{\sigma} \tag{31}$$

with some exponent σ, that within our numerical precision, does not depend on temperature (see below). So the present results for $P_1(w)$ reveal the importance of two temperatures $T_1<T_{\rm gap}$ defined as follows. The temperature T_1 is defined by

$$\mu(T_1)=1 \tag{32}$$

in Eq. (30): for $T<T_1$, the probability distribution $P_1(w)$ is divergent as $w\to 1$ [as in Fig. 2(a)], whereas for $T>T_1$, the probability distribution $P_1(w)$ vanishes at $w=1$ [see Fig. 3(b)]. Exactly at T_1, $P_1(w=1)$ remains finite [see Fig. 3(a)]. The second temperature $T_{\rm gap}$ is defined as the last temperature where $P_1(w)$ reaches $w=1$ with some exponent $\mu(T_{\rm gap})$. For $T>T_{\rm gap}$, a gap appears in Eq. (31)

$$w_0(T_{\rm gap})=1, \tag{33}$$

$$w_0(T_{\rm gap}+\epsilon)<1. \tag{34}$$

We find that $T_{\rm gap}$ is clearly above $T_1\sim 0.095$ since at $T_2\sim 0.15$, $P_1(w)$ still reaches $w=1$ with a finite slope corresponding to $\mu(T_2)\sim 2$ as shown in Fig. 3(b).

D. Probability distribution $P_2(w)$ of the second largest weight seen by a given monomer

We show in Fig. 4 the probability distribution $P_2(w)$ of the second largest weight. The main singularities of $P_2(w)$ are the divergence near $w\to 0$ and the singularity near $w\to(1/2)^-$, which is complementary to the singularity of $P_1(w)$ at the same point $w\to(1/2)^-$ [17] [see Eq. (A19) and explanations in the Appendix]. For $T=0.02$ [Fig. 4(a)], there exists an infinite slope for $P_2(w)$ and $P_1(w)$ as $w\to(1/2)^-$. For $T=T_1\sim 0.095$ [Fig. 4(b)], there exists a cusp for $P_2(w)$ and $P_1(w)$ as $w\to(1/2)^-$.

E. Probability distribution $\Pi(Y_2)$ of the parameter Y_2

The parameter Y_2 defined in Eq. (26) can reach the value $Y_2\to 1$ only if the maximal weight $w^{\max}$ also reaches $w^{\max}\to 1$. As a consequence, the probability distribution $\Pi(Y_2)$ has the same singularity near $Y_2\to 1$ as in Eq. (30)

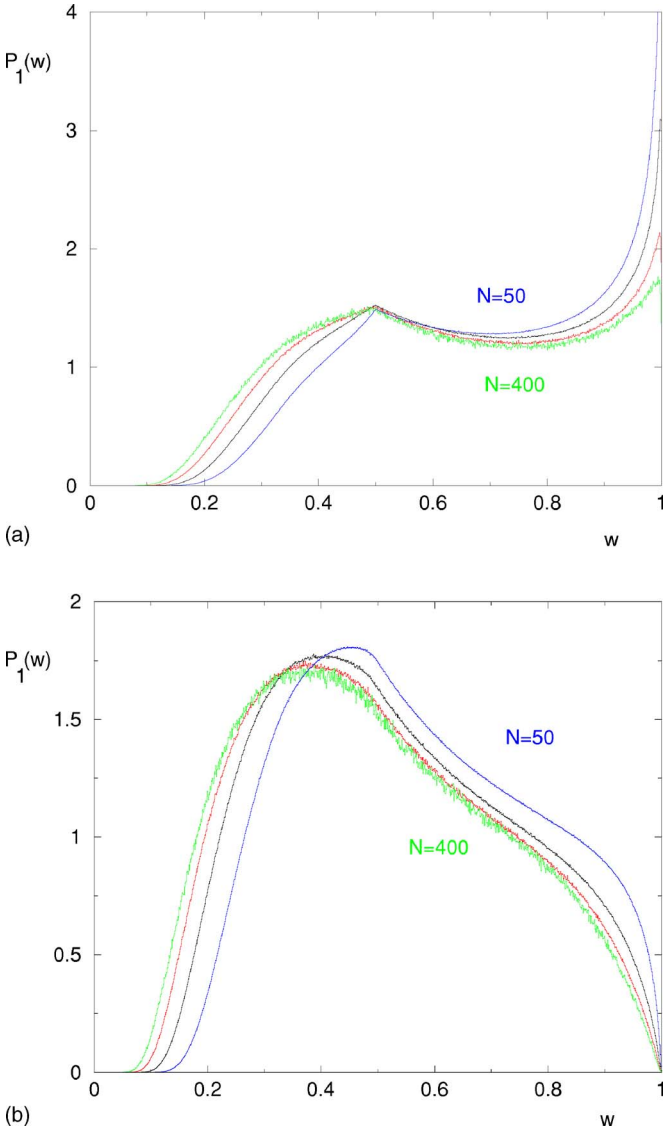


FIG. 3. (Color online) Probability distribution $P_1(w)$ of the largest weight seen by a given monomer [see Eq. (24)] (a) at $T_1\sim 0.095$ where $\mu(T_1)=1$ for $N=50,100,200,400$: $P_1(w)$ does not diverge anymore as $w\to 1$ but remains finite. (b) At $T_2=0.15$, where $\mu(T_2)\sim 2$ for $N=50,100,200,400$: the distribution $P_1(w)$ vanishes linearly as $w\to 1$.

$$\Pi(Y_2) \underset{Y_2\to 1}{\propto} (1-Y_2)^{\mu(T)-1} \tag{35}$$

for $0<T<T_{\rm gap}$ [see Fig. 5(a)], whereas a gap appear for $T>T_{\rm gap}$, as shown in Fig. 5(b).

For $T<T_{\rm gap}$, the distribution $\Pi(Y_2)$ presents the characteristic Derrida-Flyvbjerg singularities at $Y_2=1/n$ (see Appendix): in Fig. 5(a), beyond the main singularity at $Y_2=1$, the secondary singularities at $Y_2=1/2$ and at $Y_2=1/3$ are clearly visible. The distribution $\Pi(Y_2)$ is shown in Fig. 6 for the two temperatures T_1 and T_2 corresponding to $\mu(T_1)=1$ and $\mu(T_2)\sim 2$, and can be compared with the distribution $P_1(w)$ in Fig. 3.

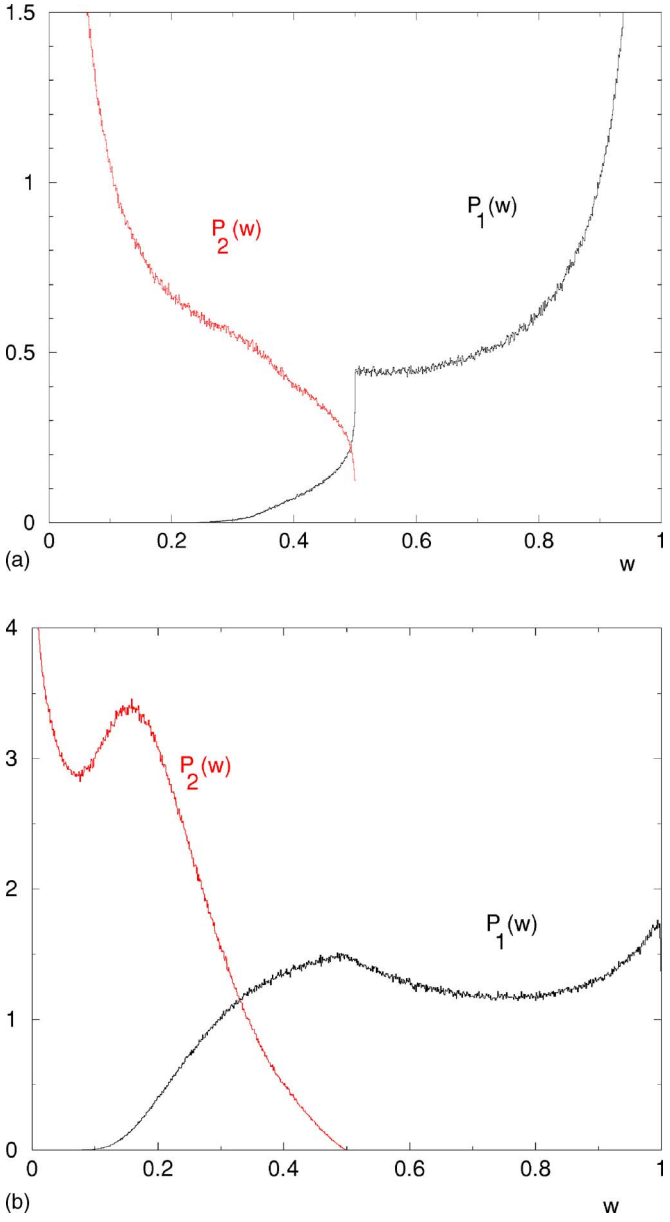


FIG. 4. (Color online) Probability distributions $P_1(w)$ and $P_2(w)$ of the two largest weights seen by a given monomer [see Eq. (24)] (a) for $T=0.02$ (here $N=200$) there exists an infinite slope for $P_2(w)$ and $P_1(w)$ as $w \to (1/2)^-$ (b) for $T_1 \sim 0.095$ (here $N=400$), there exists a cusp for $P_2(w)$ and $P_1(w)$ as $w \to (1/2)^-$.

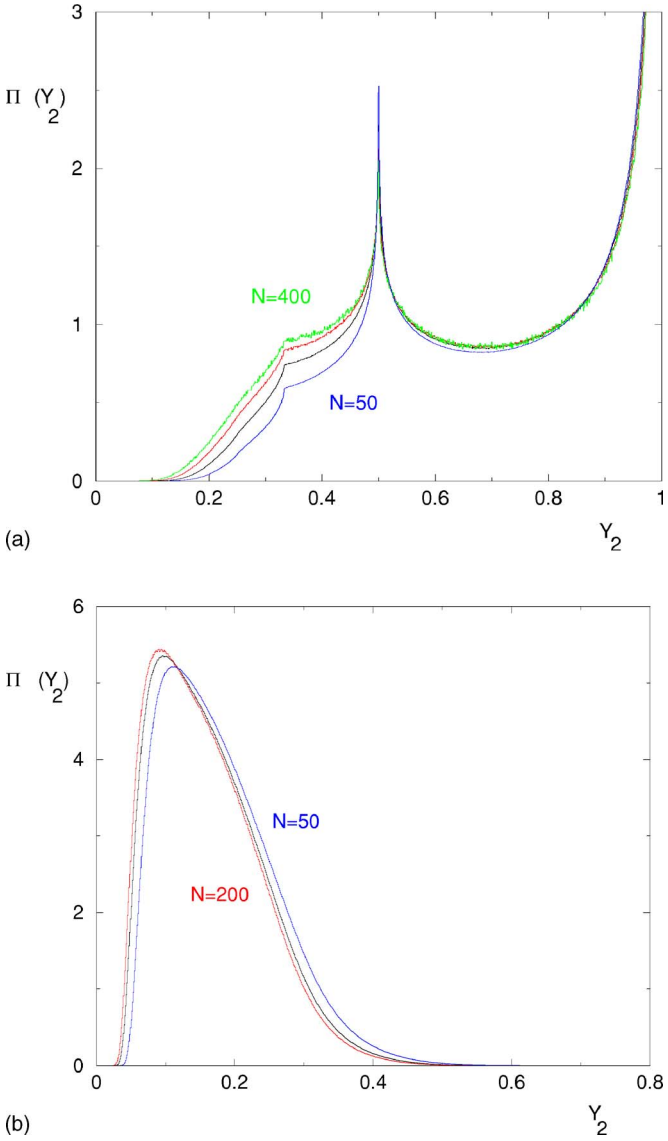


FIG. 5. (Color online) Probability distribution $\Pi(Y_2)$ of the parameter Y_2 [see Eq. (26)] (a) at $T=0.05$ (low-temperature phase) for $N=50,100,200,400$: the characteristic Derrida-Flyvbjerg singularities at $Y_2=1$, $Y_2=1/2$, and at $Y_2=1/3$ are clearly visible. (b) At $T=0.4$ (high-temperature phase) for $N=50,100,200$: the distribution $\Pi(Y_2)$ does not reach $Y_2=1$ anymore, but presents a gap.

F. Density $f(w)$

The density $f(w)$ introduced in Eq. (27) is shown on Fig. 7 at low and high temperature, respectively. By construction, this density coincides with the maximal weight distribution $P_1(w)$ for $w>1/2$, with the sum $[P_1(w)+P_2(w)]$ of the two largest weight distributions for $1/3<w<1/2$, and so on [17] [see Eq. (A19)]. As a consequence, $f(w)$ has the same singularity near $w \to 1$ as $P_1(w)$ [Eq. (30)], and the same gap [Eq. (31)] as long as $w_0(T)>1/2$. The only other singularity is near $w \to 0$ where $f(w)$ diverges in a nonintegrable manner, because in the $N \to \infty$, there is an infinite number of vanishing weights (only the product $[wf(w)]$ has to be integrable at $w=0$ as a consequence of the normalization condition of Eq. (29)).

In Fig. 8, we compare the density $f(w)$ measured in RNA with the density $f_{\text{Levy}}(w)$ [Eq. (A7)] describing the weight statistics in Lévy sums of independent variables. For small $\mu \ll 1$, the two density are rather close [Fig. 8(a)]. For larger μ, they are very different, because the density $f_{\text{Levy}}(w)$ disappears at the critical value $\mu_c=1$ [see the denominator of Eq. (A7)], whereas for RNA the density $f(w)$ exists beyond $\mu=1$.

G. Moments $\overline{Y_k(i)}$

The moments $\overline{Y_k(i)}$ [Eq. (25)] for various temperature are shown on Fig. 9 as functions of $k \leqslant 100$. The decay for large

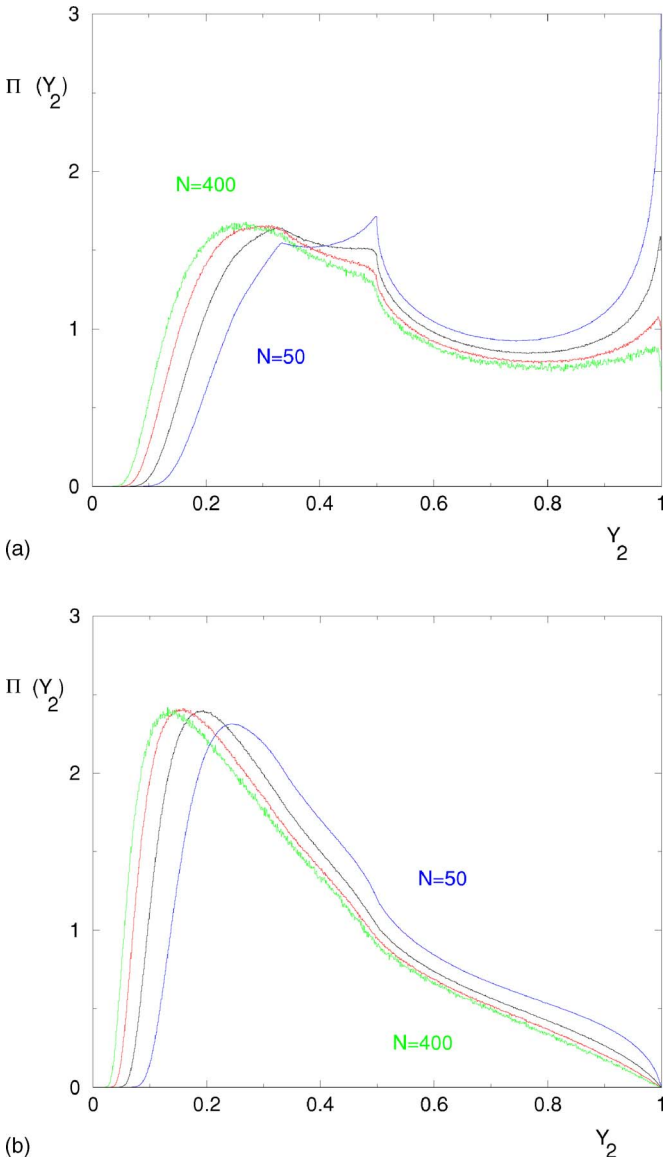


FIG. 6. (Color online) Probability distribution $\Pi(Y_2)$ of the parameter Y_2 [see Eq. (26)] (a) at $T_1 \sim 0.095$ where $\mu(T_1)=1$ for $N=50,100,200,400$: $\Pi(Y_2)$ does not diverge anymore as $Y_2 \to 1$ but remains finite and (b) at $T_2=0.15$, where $\mu(T_2) \sim 2$ for $N=50,100,200,400$: the distribution $\Pi(Y_2)$ vanishes linearly as $Y_2 \to 1$.

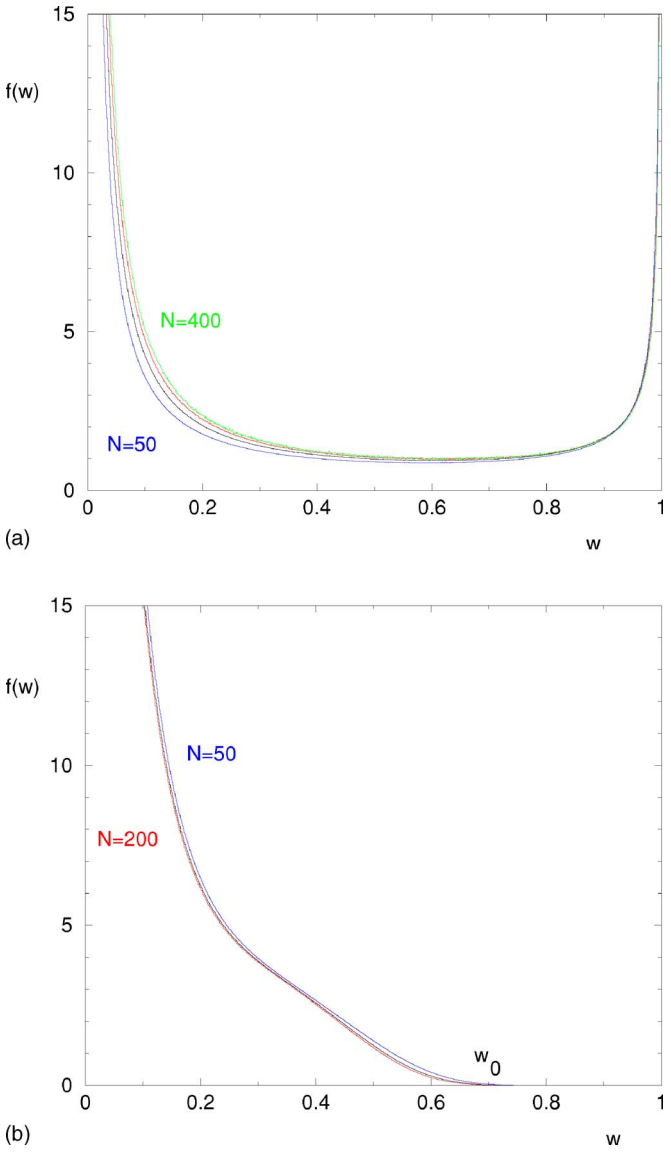


FIG. 7. (Color online) Density $f(w)$ of weights seen by a given monomer [see Eq. (27)] (a) at $T=0.05$ (low-temperature phase) for $N=50,100,200,400$. Near $w \to 1$, $f(w)$ presents the same singularity as $P_1(w)$. (b) At $T=0.4$ (high-temperature phase) for $N=50,100,200$: $f(w)$ vanishes at some value $w_0(T)<1$.

k directly reflects the behavior of the distribution of the density $f(w)$ near its maximal value, as can be seen in Eq. (28). For $0<T \leqslant T_{\text{gap}}$, where $P_1(w)$ and $f(w)$ behaves near $w \to 1$ as in Eq. (30), the decay in k follow a power-law of exponent $\mu(T)$

$$\overline{Y_k(i)} \underset{k\to\infty}{\propto} \frac{1}{k^{\mu(T)}} \text{ for } T \leqslant T_{\text{gap}}. \tag{36}$$

For $T>T_{\text{gap}}$, where there exists a gap $w_0(T)$ for $P_1(w)$, Eq. (31) also applies to $f(w)$ as long as $w_0(T)>1/2$ [since $f(w)=P_1(w)$ for $w>1/2$ as mentioned above]. This implies an exponential decay

$$\overline{Y_k(i)} \underset{k\to\infty}{\propto} \frac{[w_0(T)]^k}{k^{1+\sigma}} \text{ for } T > T_{\text{gap}}. \tag{37}$$

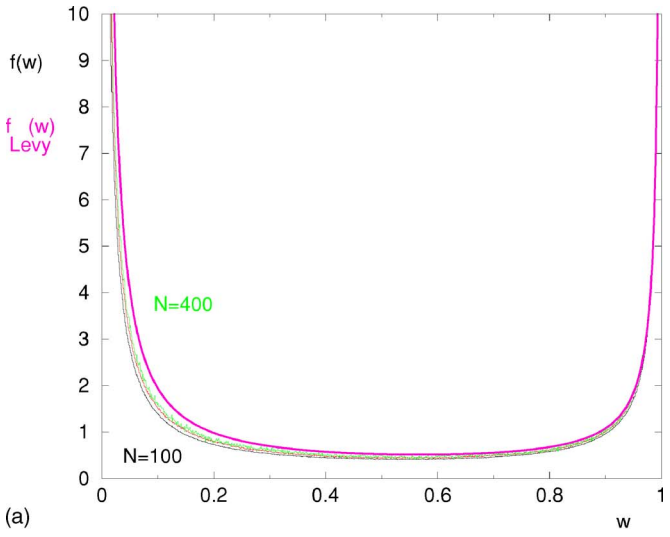
Numerically, the measure of the decay of $\overline{Y_k(i)}$ is the most convenient way to localize the temperature T_{gap} where the gap appears, and to measure the exponents. The temperature T_{gap}, where the gap appears is found to be

$$T_{\text{gap}} \sim T_2 \sim 0.15. \tag{38}$$

The exponent σ seems to be independent of T

$$\sigma \sim 0.5. \tag{39}$$

The exponent $\mu(T)$ grows with the temperature for $0<T \leqslant T_{\text{gap}}$, as shown in Fig. 9(b): the temperature T_1 where $\mu(T_1)=1$ is



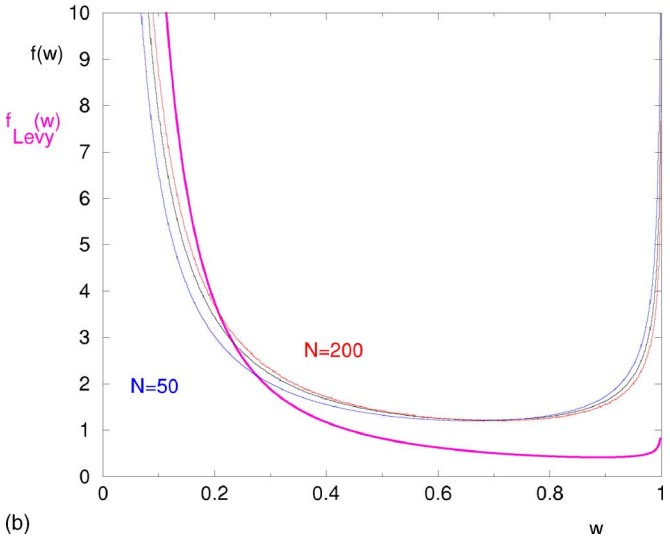
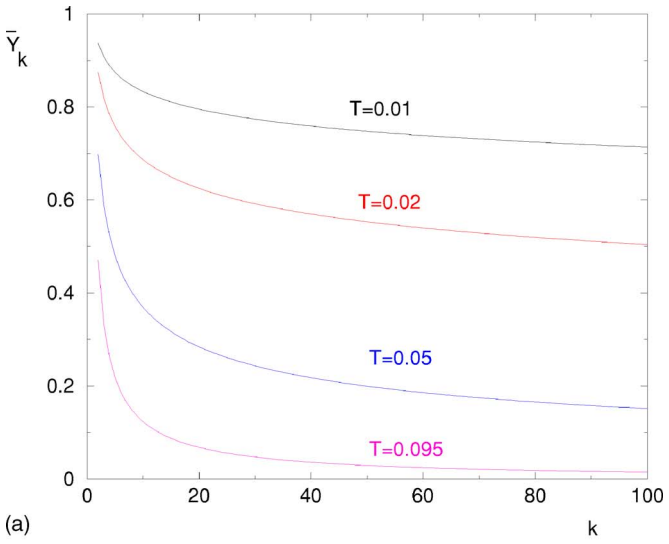


FIG. 8. (Color online) Density $f(w)$ of weights seen by a given monomer [see Eq. (27)] as compared to $f_{\mathrm{Levy}}(w)$ [Eq. (A7)] for the corresponding value of μ (thick curve) (a) at $T=0.02$, where $\mu(0.02)\sim 0.13$ (for $N=100,200,400$): $f(w)$ is rather close to the density $f_{\mathrm{Levy}}(w)$. (b) At $T=0.08$ where $\mu(0.08)\sim 0.77$ (for $N=50,100,200$): there is now a big difference between the density $f(w)$ measured for RNA and the density $f_{\mathrm{Levy}}(w)$.



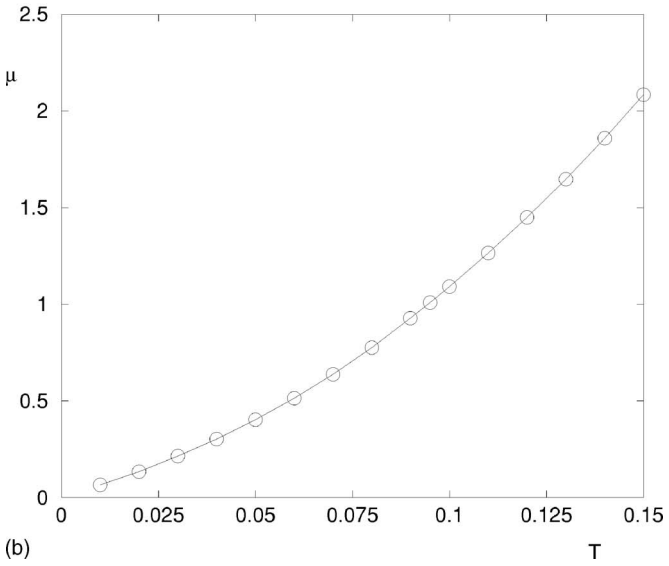


FIG. 9. (Color online) (a) Decay of the moments $\overline{Y_k(i)}$ of Eq. (36) as a function of $k\leqslant 100$ for $N=800$ and $T=0.01,0.02,0.05,0.095$. (b) Exponent $\mu(T)$ as measured from the slope of the log-log decay in the asymptotic region.

$$T_1 \sim 0.095. \tag{40}$$

In contrast with the REM (see the Appendix) where $\mu(T)=T/T_g$ is linear in the whole low temperature phase, Fig. 9(b) presents some curvature, which probably reflects the presence of some entropy. However, in the limit of very low temperature, the exponent $\mu(T)$ is linear in T and the coefficient depends on the droplet density as we now explain.

H. Droplet analysis at order T in temperature

At $T=0$, there exists a unique ground-state where each monomer (i) is paired with its ground-state partner $j_{\mathrm{g.s.}}(i)$. In the droplet analysis of disordered systems [13,18,19], various observables can be computed as first order in T in terms of the density $\rho(E=0,\lambda)dE$ of two-level excitations of energy $E\to 0$ and size λ. For instance the specific heat and the overlap are given in 1D disordered spin chains by [19,20]

$$C(T) \underset{T\to 0}{\simeq} T\frac{\pi^2}{6}\int d\lambda \rho(E=0,\lambda) + O(T^2), \tag{41}$$

$$1 - q_{EA}(T) \underset{T\to 0}{\simeq} 2T\int d\lambda \lambda \rho(E=0,\lambda) + O(T^2), \tag{42}$$

i.e., the specific heat is related to the number of excitations whereas the overlap involves the number of spins belonging to excitations. We may apply this droplet analysis to $\overline{Y_k(i)}$: the contribution at order T comes from the monomers i belonging to a droplet of energy $E\to 0$: the pair with the

ground-state partner has for weight $1/(1+e^{-\beta E})$, whereas the pair with the droplet partner has the complementary weight $e^{-\beta E}/(1+e^{-\beta E})$. Within this two-level description, one gets

$$1-\overline{Y_k(i)} \underset{T\to 0}{\simeq} \int dE \int d\lambda\lambda\rho(E,\lambda)\left[1-\left(\frac{1}{1+e^{-\beta E}}\right)^k - \left(\frac{e^{-\beta E}}{1+e^{-\beta E}}\right)^k\right] \underset{T\to 0}{\simeq} T\left(\int d\lambda\lambda\rho(E=0,\lambda)\right) I_k + O(T^2), \tag{43}$$

where the integral I_k

$$I_k = \int_{1/2}^{1} \frac{dp}{p(1-p)}[1-p^k-(1-p)^k] = \sum_{m=1}^{k-1} \frac{1}{m} \tag{44}$$

behaves logarithmically at large k

$$I_k \underset{k\to\infty}{\simeq} \ln k. \tag{45}$$

The comparison of this droplet analysis with Eq. (36) indicates that the exponent $\mu(T)$ should increase from $\mu(T=0)=0$ linearly in T with a coefficient related to the droplet density

$$\mu(T) \simeq T\left(\int d\lambda\lambda\rho(E=0,\lambda)\right) + O(T^2). \tag{46}$$

IV. STUDY OF SPATIAL PROPERTIES

In the last section, we have studied in details the statistics of the weights independently of the distance and identified two important temperatures T_1 and $T_{\rm gap}$. We now turn to the analysis various spatial properties to clarify the meaning of T_1 and $T_{\rm gap}$ for the pair length statistics.

A. Weight statistics for long-range pairs

The density $f(w)$ defined in Eq. (27) can be decomposed into l-dependent components as

$$f(w) = \sum_l f_l(w), \tag{47}$$

where $f_l(w)$ represents the density of weight of pairs of length l. At $T=\infty$, these densities are concentrated on a single l-dependent value [see Eq. (11) for $N\to\infty$]

$$[f_l(w)]_{T=\infty} \propto \delta\left(w-\frac{a}{l^{3/2}}\right), \tag{48}$$

whereas at zero temperature [see Eq. (16) for $N\to\infty$], a weight is either 0 (if the pair is not in the ground state) or 1 (if the pair belongs to the ground state)

$$[f_l(w)]_{T=0} \sim b_l\delta(w) + c_l\delta(w-1), \tag{49}$$

where the amplitude c_l of the existing weights [see Eq. (16) for $N\to\infty$] decay with l as

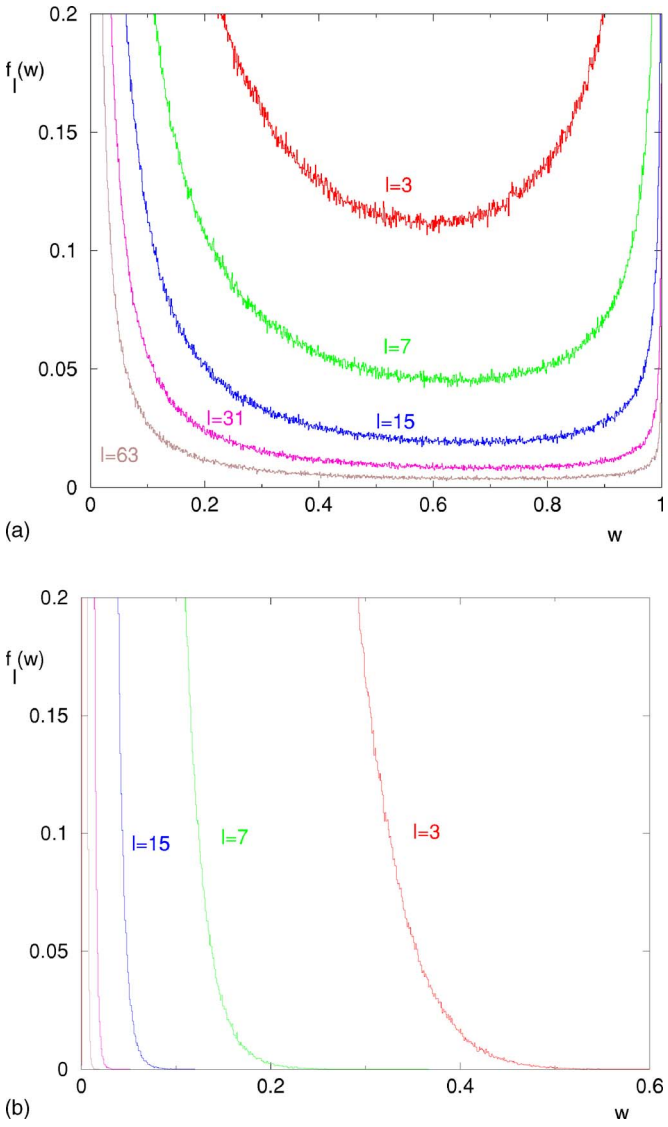


FIG. 10. (Color online) Density $f_l(w)$ of weights for pair lengths $l=3,7,15,31,63$ for size $N=200$ (a) at $T=0.05$, all curves diverge as $w\to 1$. (b) at $T=0.4$, all curves display an l-dependent gap.

$$c_l \propto \frac{1}{l^{\eta(T=0)}} = \frac{1}{l^{1.33}}. \tag{50}$$

The densities $f_l(w)$ are plotted for various lengths l at low and high temperature, respectively, in Fig. 10. At $T=0.05$ [Fig. 10(a)], all curves present divergences at $w\to 0$ and at $w\to 1$, in continuity with the two delta peaks present at $T=0$ [Eq. (49)]. At $T=0.4$ [Fig. 10(b)] all curves display an l-dependent gap, in similarity with the l-dependent delta peak of the $T=\infty$ limit [Eq. (48)].

We show in Fig. 11 the case of the two important temperatures T_1 and $T_{\rm gap}$. At $T_1\sim 0.095$ [Fig. 11(a)], all densities $f_l(w)$ still reach the value $w=1$. At $T_{\rm gap}\sim 0.15$, the densities $f_l(w)$ of small sizes l still reach the value $w=1$, whereas the densities $f_l(w)$ of large sizes l do not.

These curves suggest the following picture: (i) for $T<T_1$, all densities $f_l(w)$ diverge near $w\to 1$, so there exist

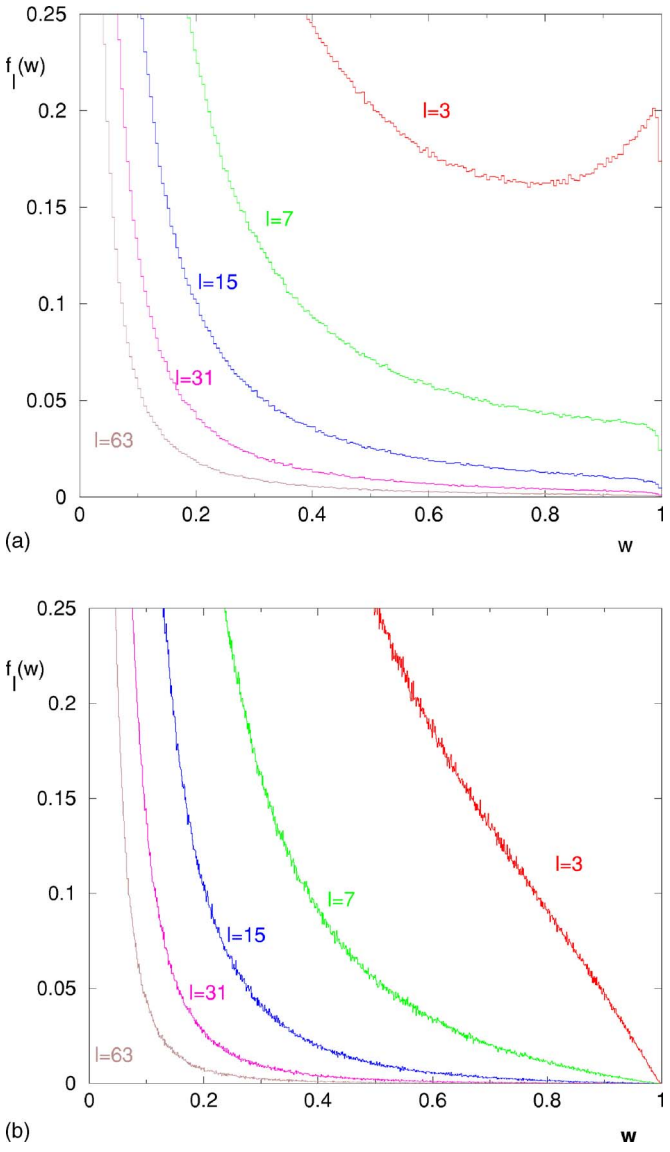


FIG. 11. (Color online) Densities $f_l(w)$ of weights for pair lengths $l=3,7,15,31,63$ for size $N=200$ (a) at $T_1 \sim 0.095$, all densities $f_l(w)$ still reach the value $w=1$ (b) at $T_{\rm gap}=0.15$, the densities $f_l(w)$ of small sizes l still reach the value $w=1$, whereas the densities $f_l(w)$ of large sizes l do not.

frozen pairs of all sizes, (ii) for $T_1 < T < T_{\rm gap}$, there exist frozen pairs, but only of finite size, (iii) for $T > T_{\rm gap}$, even short pairs are not frozen anymore. We now present various quantitative studies that confirm this scenario.

B. Statistics of the distance $l_{\rm pref}$ to the preferred partner

We now consider the probability distribution $P_N^{\rm pref}(l_{\rm pref})$ of the distance $l_{\rm pref}=|j_{\rm pref}(i)-i|$ between a base i and its preferred partner $j_{\rm pref}(i)$, i.e., the monomer $j_{\rm pref} \neq i$ having the maximal weight [Eq. (24)]. At $T=0$, this distribution coincides with the pair distribution of the ground state [Eq. (16)]

$$[P_N^{\rm pref}(l)]_{T=0} = \overline{P_{i,i+l}(T=0)} \sim \frac{1}{N^{\eta(T=0)}} \Phi\left(\frac{l}{N}\right), \tag{51}$$

whereas at $T=\infty$, the maximal weight corresponds to the nearest neighbors with $l=1$ for entropic reasons [Eqs. (11) and (13)]

$$[P_N^{\rm pref}(l)]_{T=\infty} = \delta_{l,1}. \tag{52}$$

So the first moment of this distribution

$$\overline{l^{\rm pref}} \equiv \int dl l P_N^{\rm pref}(l) \tag{53}$$

represents a correlation length that remains finite in the high temperature phase as $N \to \infty$. Since the second moment is also expected to be finite in the high temperature phase, it is convenient to define the ratio

$$B_N(T) = \frac{\overline{(l^{\rm pref})^2}}{N \overline{l^{\rm pref}}} \tag{54}$$

which converge to 0 in the high temperature phase, and to a nonzero value at criticality and in the low temperature phase. The results are shown on Fig. 12(a): the critical temperature T_c coincide with T_1

$$T_c = T_1 \sim 0.095. \tag{55}$$

The finite-size scaling of these data according to

$$B_N(T) \simeq \mathcal{B}[(T-T_c)N^{1/\nu}] \tag{56}$$

is consistent with the value

$$\nu \sim 2 \tag{57}$$

as shown on Fig. 12(b).

We now discuss the behavior of $\overline{l^{\rm pref}}$ as a function of N for various temperatures. For $T < T_c$, it grows as

$$\overline{l^{\rm pref}} \sim N^{0.67} \tag{58}$$

and the probability distribution $P_N^{\rm pref}(l)$ follows the same scaling form as in the $T=0$ limit [Eq. (51)]:

$$[P_N^{\rm pref}(l)]_{T<T_c} \sim \frac{1}{N^{1.33}} \Phi\left(\frac{l}{N}\right) \tag{59}$$

as shown in Fig. 13(a). At $T_c=T_1 \sim 0.095$, the first moment grows as

$$\overline{l^{\rm pref}} \sim N^{0.5} \tag{60}$$

and the probability distribution $P_N^{\rm pref}(l)$ follows the scaling form

$$[P_N^{\rm pref}(l)]_{T_c} \sim \frac{1}{N^{1.5}} \Psi\left(\frac{l}{N}\right) \tag{61}$$

as shown in Fig. 13(b). In Fig. 14(a), the rescaled variable $\overline{l^{\rm pref}}/N^{0.5}$ is shown as a function of T for various sizes: there is a crossing at T_c, and the data follow the finite-size scaling behavior

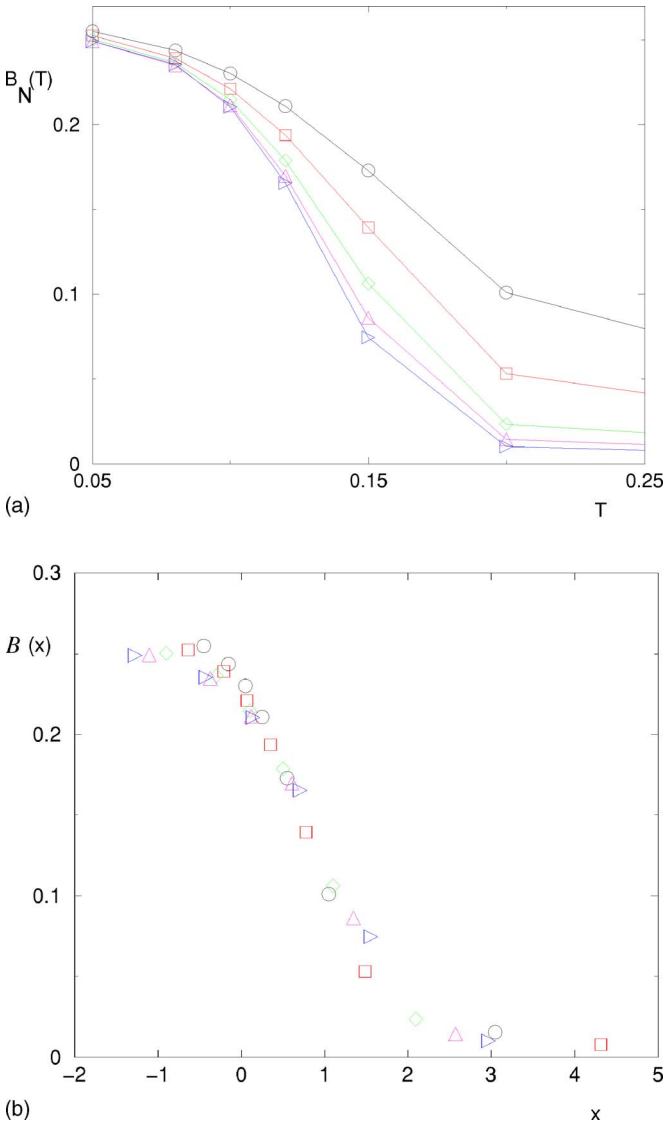


FIG. 12. (Color online) Ratio $B_N(T)$ defined in Eq. (54): (a) as a function of temperature T for sizes $N=100$ (○), $N=200$ (□), $N=400$ (◇), $N=600$ (△), $N=800$ (▷). (b) Finite size scaling of the same data in terms of the variable $x=(T-T_c)N^{1/\nu}$ [see Eq. (56)] with $T_c=0.095$ and $\nu=2$.

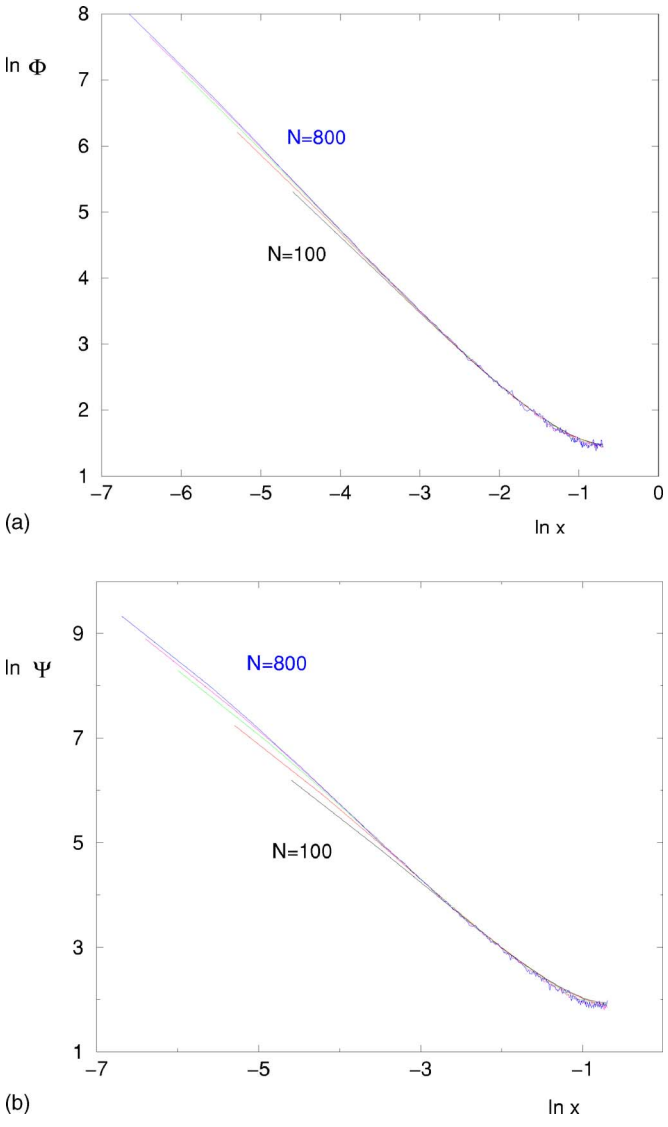


FIG. 13. (Color online) Scaling form of the probability distribution $P_N^{\mathrm{pref}}(l)$: log-log plot of $N^{\eta}P_N^{\mathrm{pref}}(l)$ in terms of $x=l/N$ (a) for $T=0.05$ (low-T phase) the rescaling is done with $\eta=1.33$ [see Eq. (59)] and (b) for $T_c\sim 0.095$ the rescaling is done with $\eta_c=1.5$ [see Eq. (61)].

$$\frac{\overline{l^{\mathrm{pref}}}}{N^{0.5}} \simeq \mathcal{L}[(T-T_c)N^{1/\nu}] \tag{62}$$

with $\nu\sim 2$ in agreement with the previous estimate of Eq. (57).

C. Pair distribution $\overline{P_{i,i+l}}$ and height scaling

We have measured the median height for each sample

$$h_{\mathrm{med}} = \frac{1}{N}\sum_k \langle h_k \rangle. \tag{63}$$

Its average over samples is directly related to the first moment of the pair distribution $\overline{P_{i,i+l}}$ [see Eq. (6)]

$$\overline{h_{\mathrm{med}}} = \sum_l l\overline{P_{i,i+l}}. \tag{64}$$

We find that in the whole low-temperature phase and at T_c, the roughness exponent is the same as at $T=0$ [Eq. (14)]

$$\overline{h_{\mathrm{med}}} \sim N^{0.67} \text{ for } 0 \leqslant T \leqslant T_c. \tag{65}$$

This is in agreement with Eq. (19) quoted from Ref. [10]. Above T_c, the crossover towards the high-temperature roughness exponent $\zeta=1/2$ [Eq. (10)] is well described by the following finite-size scaling form (see Fig. 15):

$$\frac{\overline{h_{\mathrm{med}}}}{N^{0.67}} \simeq \mathcal{H}[(T-T_c)N^{1/\nu}] \text{ for } T \geqslant T_c \tag{66}$$

with $\nu\sim 2$ in agreement with the previous estimates of Eqs. (57) and (62).

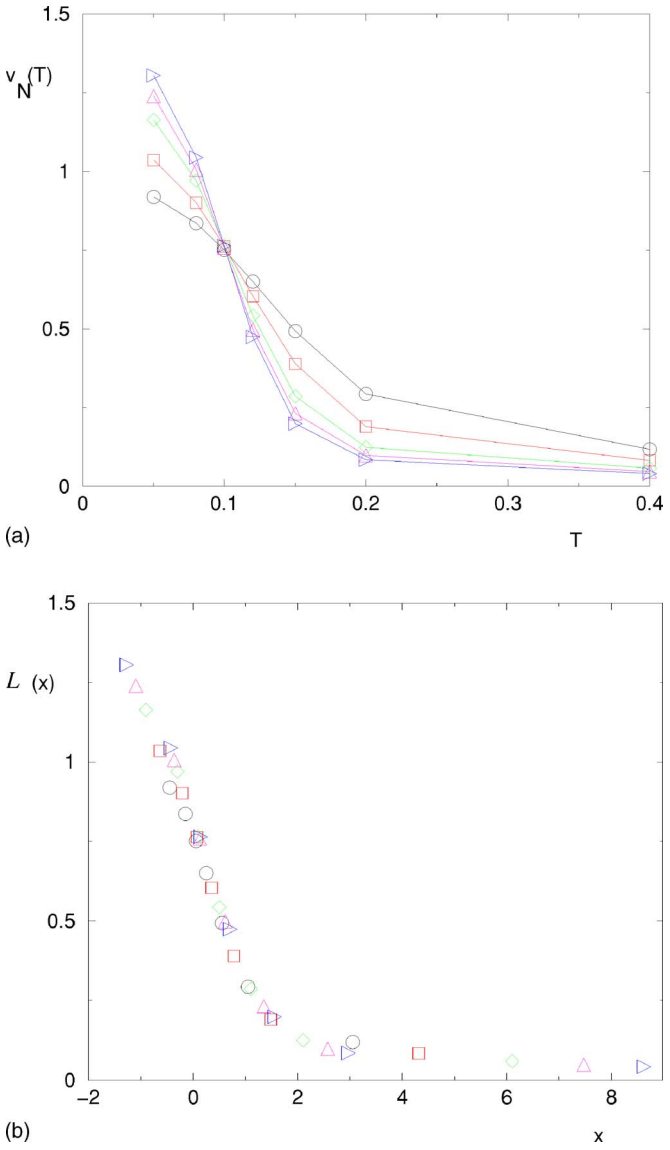


FIG. 14. (Color online) (a) Plot of $v_N(T)=\overline{l^{\mathrm{pref}}(N)}/N^{0.5}$ as a function of T for the sizes $N=100$ (○), $N=200$ (○), $N=400$ (◇), $N=600$ (△), $N=800$ (▷). (b) Finite size scaling of the same data in terms of the variable $x=(T-T_c)N^{1/\nu}$ [see Eq. (62)] with $T_c=0.095$ and $\nu=2$.

Accordingly, we find that the pair distribution $\overline{P_{i,i+l}}$ follows the $T=0$ finite-size scaling of Eq. (16) in the whole low-temperature phase and also at T_c

$$\overline{P_{i,i+l}(T_c)} \sim \frac{1}{N^{1.33}}\Phi\left(\frac{l}{N}\right). \tag{67}$$

D. Overlap $\overline{P^2_{i,i+l}}$ of large pairs l

We have also measured the overlap $\overline{P^2_{i,i+l}}$ of large pairs. We find that the first moment scales as

$$\int dl l\overline{P^2_{i,i+l}} \sim N^{0.67} \text{ for } 0 \leqslant T < T_c, \tag{68}$$

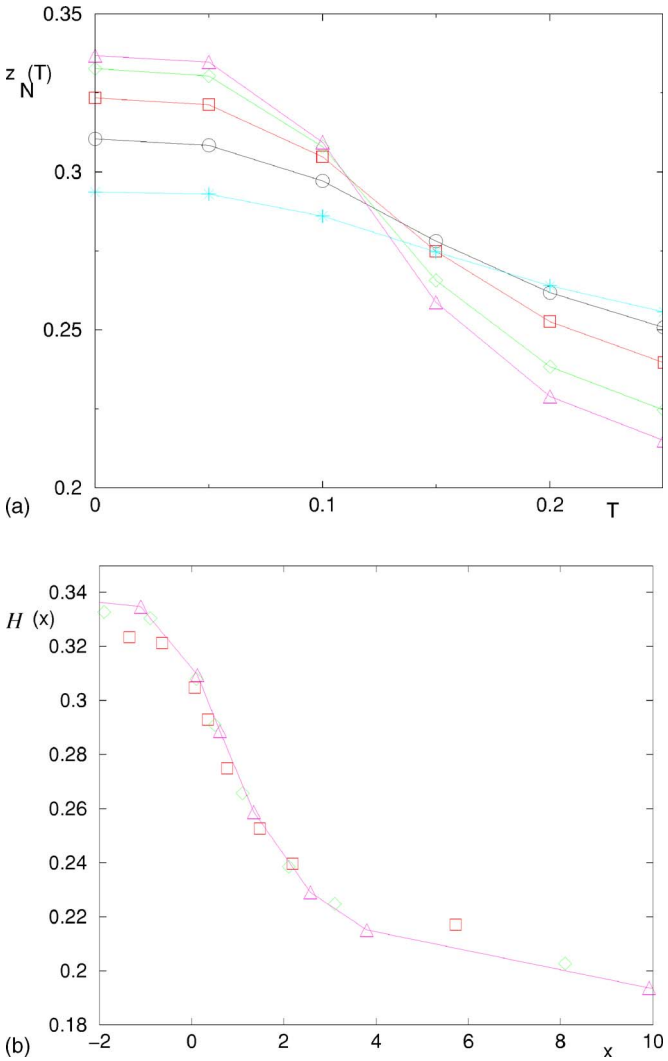


FIG. 15. (Color online) Height scaling: (a) the curves $z_N(T)=\overline{h_{\mathrm{med}}}/N^{0.67}$ of various sizes $N=50$ (*), $N=100$ (○), $N=200$ (□), $N=400$ (◇), $N=600$ (△) present crossings shifting towards T_c. (b) Finite size scaling of the same data in terms of the variable $x=(T-T_c)N^{1/\nu}$ [see Eq. (66)] with $T_c=0.095$ and $\nu=2$.

$$\int dl l\overline{P^2_{i,i+l}} \sim N^{0.5} \text{ for } T = T_c, \tag{69}$$

$$\int dl l\overline{P^2_{i,i+l}} \sim cte \text{ for } T > T_c. \tag{70}$$

The scaling exactly at T_c is distinct from the low-temperature phase in disagreement with Eq. (20) quoted from Ref. [10], but coincides with the scaling found above for $\overline{l^{\mathrm{pref}}}$ [Eq. (60)]. It is thus convenient to define the ratio

$$R_N(T) = \frac{\int dl l\overline{P^2_{i,i+l}}}{\int dl l\overline{P_{i,i+l}}} \tag{71}$$

which converge to 0 in the high temperature phase, and to a nonzero value in the low temperature phase. Exactly at T_c, it

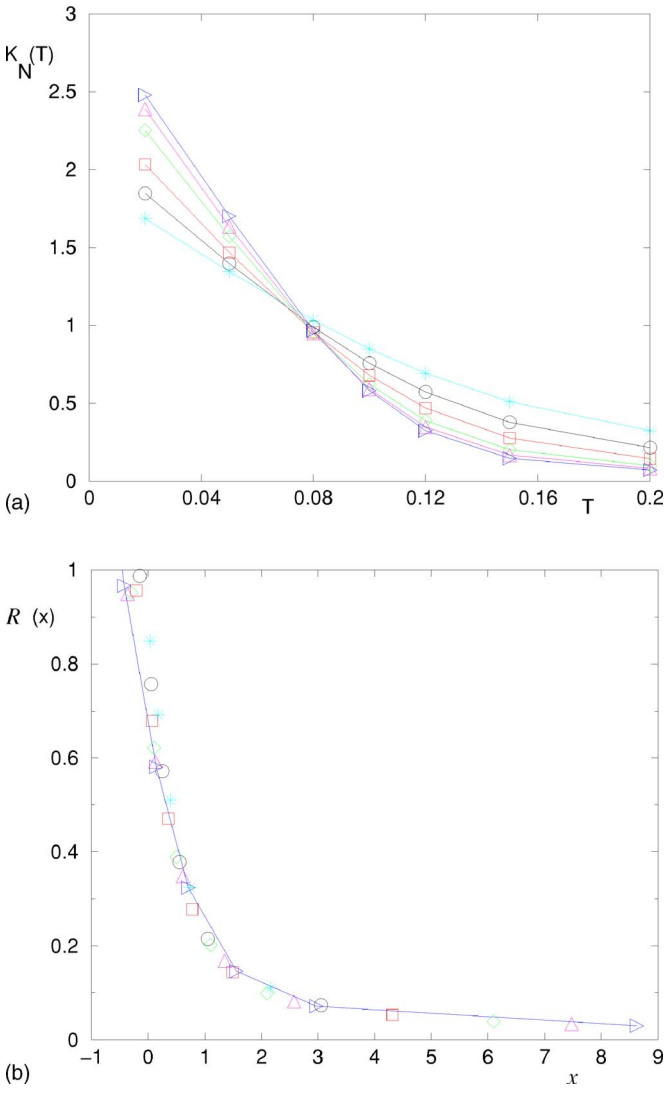


FIG. 16. (Color online) Critical behavior of the ratio $R_N(T)$ defined in Eq. (71): (a) curves $K_N(T)=N^{0.17}R_N(T)$ for the sizes $N=50$ (*) $N=100$ (○), $N=200$ (□), $N=400$ (◇), $N=600$ (△), $N=800$ (▷). (b) Finite size scaling of the same data in terms of the variable $x=(T-T_c)N^{1/\nu}$ [see Eq. (72)] with $T_c=0.095$ and $\nu=2$.

is expected to decay as $N^{0.5}/N^{0.67}=N^{-0.17}$. In Fig. 16(a), the curves $K_N(T)=N^{0.17}R_N(T)$ present crossings that shift regularly towards T_c. The finite-size scaling of these data according to

$$K_N(T)=N^{0.17}R_N(T)\simeq\mathcal{R}[(T-T_c)N^{1/\nu}] \tag{72}$$

with $T_c=0.095$ and $\nu=2$ is shown on Fig. 16(b).

V. SUMMARY AND CONCLUSIONS

In this paper, we have analyzed the freezing transition of random RNA secondary structures via the statistics of the pairing weights seen by a given monomer. In analogy with Lévy sums and Derrida's random energy model [16,17], we have numerically computed the probability distributions $P_1(w)$ of the maximal weight, the probability distribution $\Pi(Y_2)$ of the parameter $Y_2(i)=\Sigma_j w_i^2(j)$, as well as the average values of the moments $Y_k(i)=\Sigma_j w_i^k(j)$. We have found two important temperatures $T_c<T_{\rm gap}$. For $T>T_{\rm gap}$, the distribution $P_1(w)$ and $\Pi(Y_2)$ have a gap and, accordingly, the moments $\overline{Y_k(i)}$ decay exponentially in k. For $T<T_{\rm gap}$, these moments decay with a power-law $\overline{Y_k(i)}\sim 1/k^{\mu(T)}$, and the distributions $P_1(w)$ and $\Pi(Y_2)$ present the characteristic Derrida-Flyvbjerg singularities at $w=1/n$ and $Y_2=1/n$ for $n=1,2,\ldots$. The most important singularities occur at $w=1$ with $P_1(w)\sim(1-w)^{\mu(T)-1}$ and $\Pi(Y_2)\sim(1-Y_2)^{\mu(T)-1}$. The exponent $\mu(T)$ increases with the temperature from the value $\mu(T=0)=0$ up to $\mu(T_{\rm gap})\sim 2$. The study of spatial properties indicates that the critical temperature T_c where the large-scale roughness exponent changes from the low temperature value $\zeta\sim 0.67$ to the high temperature value $\zeta\sim 0.5$ corresponds to the exponent $\mu(T_c)=1$. The final picture is thus as follows: (i) for $T<T_c$, there exist frozen pairs of all sizes, (ii) for $T_c<T<T_{\rm gap}$, there exist frozen pairs, but they are of finite-size, (iii) for $T>T_{\rm gap}$, even short pairs are not frozen anymore. Finally, the finite-size scaling of various data are consistent with a correlation length exponent $\nu\simeq 2$ that saturates the general bound $\nu\geqslant 2/d=2$ of [14] for phase transitions in disordered systems.

In conclusion, the numerical study of the weight statistics appears as an interesting tool to clarify the nature of low temperature phases existing in disordered systems. In particular, we have shown that the frozen phase is characterized by a temperature-dependent exponent $\mu(T)$ that governs the broadening of the delta peak existing at $w=1$ at $T=0$. We intend to study in a similar way other disordered models [21].

APPENDIX: REMINDER ON LÉVY SUMS, THE RANDOM ENERGY MODEL, AND DERRIDA-FLYVBJERG SINGULARITIES

1. Lévy sums when the first moment is infinite

In this section, we recall some results on the weight statistics [16] for the case of Lévy sums

$$S_N=\sum_{i=1}^{N}x_i \tag{A1}$$

of N positive independent variables $(x_1,\ldots,x_N)$ distributed with a probability distribution that decays algebraically

$$\rho(x)\underset{x\to+\infty}{\simeq}\frac{A}{x^{1+\mu}} \tag{A2}$$

with $0<\mu<1$, i.e., when the first moment diverges $\langle x\rangle=+\infty$. The sum S_N then grows as $N^{1/\mu}$, and the rescaled variable is distributed with a stable Lévy distribution [22]. Another important property is that the maximal variable $x_{\max}(N)$ among the N variables $(x_1,\ldots,x_N)$ is also of order $N^{1/\mu}$, i.e., the sum S_N is actually dominated by the few biggest terms [16,22]. To quantify this effect, it is convenient to introduce the weights

$$w_i = \frac{x_i}{S_N} \tag{A3}$$

and their moments

$$Y_k = \sum_{i=1}^{N} w_i^k. \tag{A4}$$

In particular, their averaged values in the limit $N \to \infty$ are finite for $0 < \mu < 1$ and reads [16]

$$\overline{Y_k}^{\text{Levy}} = \frac{\Gamma(k-\mu)}{\Gamma(k)\Gamma(1-\mu)}. \tag{A5}$$

The density $f(w)$ giving rise to these moments

$$\overline{Y_k}^{\text{Levy}} = \int_0^1 dw w^k f(w) \tag{A6}$$

reads

$$f_{\text{Levy}}(w) = \frac{w^{-1-\mu}(1-w)^{\mu-1}}{\Gamma(\mu)\Gamma(1-\mu)} \tag{A7}$$

and represents the averaged number of terms of weight w. This density is nonintegrable as $w \to 0$, because in the limit $N \to \infty$, the number of terms of vanishing weights diverges. The normalization corresponds to

$$\overline{Y_{k=1}}^{\text{Levy}} = \int_0^1 dw w f(w) = 1. \tag{A8}$$

More generally, correlations functions between Y_k can also be computed [16], and the joint density of K weights reads [17]

$$f(w_1, \ldots, w_K) = \frac{\mu^{K-1}\Gamma(K)}{\Gamma^K(1-\mu)\Gamma(K\mu)} \left(\prod_{i=1}^{K} w_i^{-1-\mu} \right) \times \left(1 - \sum_{i=1}^{K} w_i \right)^{K\mu-1}. \tag{A9}$$

2. Reminder on the random energy model

The random energy model (REM) introduced in the context of spin glasses [23] is defined by the partition function of N spins

$$Z_N = \sum_{\alpha=1}^{2^N} e^{-\beta E_\alpha}, \tag{A10}$$

where the (2^N) energy levels are independent random variables drawn from the Gaussian distribution

$$P_N(E) = \frac{1}{\sqrt{\pi N}} e^{-E^2/N}. \tag{A11}$$

It turns out that the low temperature phase $0 < T < T_g$ of the REM [16,24] is in direct correspondence with Lévy sums of index $0 < \mu = \frac{T}{T_g} < 1$: the weights

$$w_\alpha = \frac{e^{-\beta E_\alpha}}{Z_N} \tag{A12}$$

have exactly the same moments Y_k [Eq. (A5)] and the same density $f(w)$ [Eq. (A7)]. The explanation is that the lowest energy in the REM are distributed exponentially

$$P_{\text{extremal}}(E) \underset{E \to -\infty}{\simeq} e^{\gamma E}. \tag{A13}$$

This exponential form that corresponds to the tail of the Gumbel distribution for extreme-value statistics [25,26], immediately yields that the Boltzmann weight $x = e^{-\beta E}$ has a distribution that decays algebraically [Eq. (A2)] with exponent

$$\mu = T\gamma. \tag{A14}$$

In the REM, the coefficient γ in the exponential [Eq. (A13)] is $\gamma = 1/T_g$.

The link with the thermodynamics is that the entropy S_N remains finite as N grows and is given in terms of the weights by [27]

$$\begin{aligned} S_N(T < T_g) &= -\sum_{i=1}^{N} w_i \ln(w_i) \\ &= -\left[\partial_k \sum_{i=1}^{N} w_i^k \right]_{k \to 1} \\ &= -[\partial_k Y_k]_{k \to 1} \\ &= \Gamma'(1) - \frac{\Gamma'[1-\mu(T)]}{\Gamma[1-\mu(T)]} \end{aligned} \tag{A15}$$

and the corresponding specific heat $C_N(T < T_g) = T\partial_T S_N(T)$ then coincides with the finite-size result computed in Ref. [23]. So the entropy per spin S_N/N and the specific heat C_N/N vanish as $N \to \infty$ in the whole low-T phase. In the critical region $T \to T_g^-$, the finite-size scaling behavior is

$$\frac{C_N(T)}{N} \propto \frac{1}{N(T_g - T)^2}. \tag{A16}$$

As a final remark, let us mention that in the mean-field Sherrington-Kirkpatrick (SK) model of spin-glasses, the same expressions of Y_k [Eq. (A5)] also appear [12,28], but with a different interpretation: the weights are those of the pure states. As a consequence, the parameter $\mu(T)$, which is a complicated function of the temperature, vanishes at the transition $\mu(T_c) = 0$ (only one pure state in the high temperature state) and grows as T is lowered towards $\mu(T=0)$ of order 0.5 [29]. This is in contrast with the REM model where $\mu(T) = T/T_g$ grows with the temperature from $\mu(T=0) = 0$ (only one ground state) to $\mu(T_g) = 1$ at the transition, where the number of important microscopic states is not finite anymore. Nevertheless, the expression [Eq. (A5)] for the weights of pure states means that the free-energy f of pure states in the SK model is distributed exponentially

$$P(f) \underset{f \to -\infty}{\simeq} e^{\gamma(T) f} \tag{A17}$$

with a parameter $\gamma(T) = \mu(T)/T$.

3. Derrida-Flyvbjerg singularities

In Ref. [17], the statistics of the weights w_i normalized to

$$\sum_i w_i = 1 \tag{A18}$$

for Lévy sums with $0<\mu<1$ or equivalently of the REM or SK model have been studied in detail. In particular, the probability distributions $P_1(w)$ of the maximal weight $w^{\max}=\max_i[w_i]$, $P_2(w)$ of the second maximal weight, and $\Pi(Y_2)$ of the parameter $Y_2=\Sigma_i w_i^2$ present singularities at $w=1/n$ and $Y_2=1/n$ for $n=1,2\ldots$: this shows that all the weight is concentrated on a few terms.

The origin of these singularities is that the density of weights given in Eq. (A7) satisfy [17]

$$f(w) = P_1(w) \text{ for } \frac{1}{2} < w < 1,$$

$$f(w) = P_1(w) + P_2(w) \text{ for } \frac{1}{3} < w < \frac{1}{2},$$

$$f(w) = P_1(w) + \cdots + P_n(w) \text{ for } \frac{1}{n+1} < w < \frac{1}{n}. \tag{A19}$$

For $w\to 1$, $P_1(w)$ thus presents the same singular behavior as $f(w)$ of Eq. (A7)

$$P_1(w=1-\epsilon) \underset{\epsilon\to 0}{\propto} \epsilon^{\mu-1}. \tag{A20}$$

For $w\to 1/2$, the singularity of $P_1(w)$ comes from the two different expressions of Eqs. (A19)

$$\begin{aligned} &P_1\left(w=\frac{1}{2}+\epsilon\right) - P_1\left(w=\frac{1}{2}-\epsilon\right) \\ &= f\left(w=\frac{1}{2}+\epsilon\right) - \left[f\left(w=\frac{1}{2}-\epsilon\right) - P_2\left(\frac{1}{2}-\epsilon\right)\right] \\ &\underset{\epsilon\to 0}{\propto} P_2\left(\frac{1}{2}-\epsilon\right). \end{aligned} \tag{A21}$$

For $1/3<w_2<w_1$, the joint probability that the two largest weights are w_1 and w_2 is given by Eq. (A9) for $K=2$

$$f(w_1,w_2) = \frac{\mu}{\Gamma^2(1-\mu)\Gamma(2\mu)} w_1^{-1-\mu} w_2^{-1-\mu} (1-w_1-w_2)^{2\mu-1} \tag{A22}$$

and thus $P_2(w_2)$ reads for $1/3<w_2<1/2$

$$\begin{aligned} P_2(w_2) &= \int_{w_2}^{1-w_2} dw_1 f(w_1,w_2) = \frac{\mu}{\Gamma^2(1-\mu)\Gamma(2\mu)} w_2^{-1-\mu} \\ &\times \int_{w_2}^{1-w_2} dw_1 w_1^{-1-\mu} (1-w_1-w_2)^{2\mu-1}. \end{aligned} \tag{A23}$$

The singularity near $w_2\to 1/2$ is thus of order

$$P_2\left(w_2=\frac{1}{2}-\epsilon\right) \propto \int_{1/2-\epsilon}^{1/2+\epsilon} dw_1 \left(\frac{1}{2}+\epsilon-w_1\right)^{2\mu-1} \propto \epsilon^{2\mu}. \tag{A24}$$

So for $w_1, w_2\to 1/2$, $P_1(w_1)$ and $P_2(w_2)$ have a singularity of order $\epsilon^{2\mu}$, i.e., there is no divergence, in contrast with the singularity near $w\to 1$, but there is an infinite slope for $0<\mu<1/2$. More generally, the singularities of $P_1(w_1)$ at $w_1=1/n-\epsilon$ are weaker and weaker as n grows according to [17]

$$P_1\left(w_1=\frac{1}{n}-\epsilon\right) \propto \epsilon^{2\mu-2+n}. \tag{A25}$$

Similarly, for the distribution $\Pi(Y_2)$, the singularities are given by [17]

$$\Pi\left(Y_2=\frac{1}{n}-\epsilon\right) \propto \epsilon^{n(\mu+1/2)-3/2}. \tag{A26}$$

In particular, near $Y_2\to 1$, $\Pi(Y_2)$ exhibits the same divergence as $P_1(w\to 1)$

$$\Pi(Y_2=1-\epsilon) \propto \epsilon^{\mu-1} \tag{A27}$$

and near $Y_2\to 1/2$, the singularity

$$\Pi\left(Y_2=\frac{1}{2}-\epsilon\right) \propto \epsilon^{2\mu-1/2} \tag{A28}$$

corresponds to an infinite slope for $0<\mu<\frac{3}{4}$. We refer the reader to Ref. [17] for more details. The shapes of $P_1(w)$, $P_2(w)$, and $\Pi(Y_2)$ can be found for $\mu=0.3$ and $\mu=0.1$ on Figs. 3 and 4 of Ref. [17] respectively. Similar Derrida-Flyvbjerg singularities describe above for the case of Lévy sums or spin-glasses (REM or SK), actually occur in many other contexts, such as randomly broken objects [17,30], in population genetics [31–33], in random walk excursions or loops [16,34,35].

[1] P. G. Higgs, Q. Rev. Biophys. **33**, 199 (2000).
[2] P. Schuster, Rep. Prog. Phys. **69**, 1419 (2006).
[3] P. G. Higgs, Phys. Rev. Lett. **76**, 704 (1996).
[4] S. R. Morgan and P. G. Higgs, J. Phys. A **31**, 3153 (1998).
[5] F. Krzakala, M. Mézard, and M. Müller, Europhys. Lett. **57**, 752 (2002); M. Müller, F. Krzakala, and M. Mézard, Eur. Phys. J. E **9**, 67 (2002); Markus Müller, Ph.D. thesis, Universite Paris 11, Orsay, 2003. Available from http://www.physics.harvard.edu/markusm/
[6] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999);

Europhys. Lett. **59**, 903 (2002); Phys. Rev. E **65**, 031903 (2002).
[7] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000); A. K. Hartmann, *ibid.* **86**, 1382 (2001).
[8] E. Marinari, A. Pagnani, and F. Ricci-Tersenghi, Phys. Rev. E **65**, 041919 (2002).
[9] B. Burghardt and A. K. Hartmann, Phys. Rev. E **71**, 021913 (2005).
[10] M. Lässig and K. J. Wiese, Phys. Rev. Lett. **96**, 228101 (2006); F. David and K. J. Wiese, e-print q-bio.BM/0607044.
[11] S. Hui and L.-H. Tang, e-print q-bio.BM/0608020.
[12] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore, 1987).
[13] D. S. Fisher and D. A. Huse, Phys. Rev. B **38**, 386 (1988).
[14] J. T. Chayes, L. Chayes, D. S. Fisher, and T. Spencer, Phys. Rev. Lett. **57**, 2999 (1986).
[15] B. Derrida, R. B. Griffiths, and P. G. Higgs, Europhys. Lett. **18**, 361 (1992).
[16] B. Derrida, in *On Three Levels*, edited by M. Fannes *et al.* (Plenum Press, New York, 1994).
[17] B. Derrida and H. Flyvbjerg, J. Phys. A **20**, 5273 (1987).
[18] D. S. Fisher and D. A. Huse, Phys. Rev. B **43**, 10728 (1991).
[19] C. Monthus and P. Le Doussal, Eur. Phys. J. B **41**, 535 (2004).
[20] C. Monthus and T. Garel, Phys. Rev. B **71**, 094436 (2005).
[21] C. Monthus and T. Garel, e-print cond-mat/0701699; e-print cond-mat/0702131.
[22] J. P. Bouchaud and A. Georges, Phys. Rep. **195**, 127 (1990).
[23] B. Derrida, Phys. Rev. B **24**, 2613 (1981).
[24] B. Derrida and G. Toulouse, J. Phys. (France) Lett. **46**, L223 (1985).
[25] E. J. Gumbel, *Statistics of Extreme* (Columbia University Press, New York, 1958); J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (Krieger, Malabar, FL, 1987).
[26] J. P. Bouchaud and M. Mézard, J. Phys. A **30**, 7997 (1997).
[27] D. J. Gross and M. Mézard, Nucl. Phys. B **240**, 431 (1984).
[28] M. Mézard, G. Parisi, and M. A. Virasoro, J. Phys. (France) Lett. **46**, L217 (1985).
[29] J. Vannimenus, G. Toulouse, and G. Parisi, J. Phys. (France) **42**, 565 (1981); A. Crisanti, T. Rizzo, and T. Temesvari, Eur. Phys. J. B **33**, 203 (2003).
[30] P. L. Krapivsky, I. Grosse, and E. Ben-Naim, Phys. Rev. E **61**, R993 (2000).
[31] P. G. Higgs, Phys. Rev. E **51**, 95 (1995).
[32] B. Derrida and B. Jung-Muller, J. Stat. Phys. **94**, 277 (1999).
[33] B. Derrida, in *Nonlinear Cooperative Phenomena in Biological Systems*, edited by L. Matsson (World Scientific, Singapore, 1998), pp. 216–226.
[34] L. Frachebourg, I. Ispolatov, and P. L. Krapivsky, Phys. Rev. E **52**, R5727 (1995).
[35] D. Ertas and Y. Kantor, Phys. Rev. E **53**, 846 (1996); S. Wolfling and Y. Kantor, Eur. Phys. J. B **12**, 569 (1999).