Phase transition in random planar diagrams and RNA-type matching

Andrey Y. Lokhov,[1] Olga V. Valba,[1,2] Mikhail V. Tamm,[3] and Sergei K. Nechaev[1,4]

[1]*Université Paris-Sud/CNRS, LPTMS, UMR8626, Bât. 100, 91405 Orsay, France*
[2]*Moscow Institute of Physics and Technology, 141700, Dolgoprudny, Russia*
[3]*Physics Department, Moscow State University, 119992, Moscow, Russia*
[4]*P. N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991, Moscow, Russia*

We study the planar matching problem, defined by a symmetric random matrix with independent identically distributed entries, taking values zero and one. We show that the existence of a perfect planar matching structure is possible only above a certain critical density, p_c, of allowed contacts (i.e., of ones). Using a formulation of the problem in terms of Dyck paths and a matrix model of planar contact structures, we provide an analytical estimation for the value of the transition point, p_c, in the thermodynamic limit. This estimation is close to the critical value, $p_c \approx 0.379$, obtained in numerical simulations based on an exact dynamical programming algorithm. We characterize the corresponding critical behavior of the model and discuss the relation of the perfect-imperfect matching transition to the known molten-glass transition in the context of random RNA secondary structure formation. In particular, we provide strong evidence supporting the conjecture that the molten-glass transition at $T = 0$ occurs at p_c.

I. INTRODUCTION

In this paper the combinatorial problem of complete planar matching is considered. It can be formulated as follows. Take L points $i = 1, \ldots, L$ on a line, and define a symmetric $L \times L$ random matrix A containing ones or zeroes. We are looking for a set of $L/2$ links between pairs of points allowed by the entries A_{ij} (each point is involved in one link only) such that these links form a planar diagram (see Fig. 1). This problem can be thought of as a constraint satisfaction problem (CSP) characterized by a certain distribution $P(A_{ij})$ on the entries of the matrix A. If at least one such set of links exists, we say that the problem is satisfiable, and we refer to the solution as to the *perfect matching* configuration.

The Bernoulli model that we study in this paper is defined as follows: $A_{ij} = A_{ji}$ are independent identically distributed random variables, equal to one with probability p for any $i \neq j$, and equal to zero otherwise. This sets the uniform distribution on the entries of the matrix A ($A_{ij} = A_{ji}$):

$$\text{Prob}(A_{ij}) = p\delta(A_{ij} - 1) + (1 - p)\delta(A_{ij}), \qquad (1)$$

where $\delta(x) = 1$ for $x = 0$, and $\delta(x) = 0$ otherwise.

The matrix A can be regarded as an adjacency matrix of a random Erdös-Rényi graph $G(A)$ without self-connections. In what follows, we describe a phase transition, typical for CSP's. A well-known example of such transition is the SAT-UNSAT problem [1]. As the number of constraints per node, imposed by the matrix A, is below some certain critical value, p_c, the problem is satisfiable, while above p_c it becomes unsatisfiable in the large L limit. In other words, there is a critical value of the bond formation probability, p_c, such that for any large ($L \gg 1$) sample of the matrix A, for $p > p_c$ it is always possible to find at least one "gapless" planar diagram, which involves in its formation almost all vertices and only $\sim o(L)$ vertices are missing, while for $p < p_{cr}$ a finite fraction of missing vertices of order $\sim O(L)$ exists; see Fig. 1(a) and 1(b).

To the best of our knowledge, in the context of random matrix theory, this transition has never been discussed in the literature. The planar diagrams play an important role in various branches of theoretical physics, such as matrix and gauge theories [2], many-body condensed matter physics [3], quantum spin chains [4], and random matrix theory [5]. Since we do not assume the sparsity of the matrix A, the spectral techniques developed for the sparse random matrices [6–8] seem to be not applicable in the present problem.

Besides the mathematical context, this problem has a straightforward application to finding the optimal secondary structure in RNA molecules [9–20]. The secondary RNA structure consists of a set of saturating reversible chemical bonds between monomers. These bonds correspond to links represented by planar diagrams. In the RNA context, the matrix A is a function of a frozen disorder in the sequence of monomers (nucleotides). Recently the existence of a matching transition as a function of the number of different monomer types (the nucleotide alphabet size) has been demonstrated in Ref. [21]. The main contribution of our current work is twofold. On one hand, it consists in the description of the "perfect-imperfect" phase transition and the determination of the threshold value p_c, using both analytic methods and numerical computations. On the other hand, we treat the relation between a zero-temperature perfect-imperfect matching transition in random planar diagrams, and a temperature-dependent molten-glass transition in random RNAs, widely discussed in the literature [13,15,22,23]. We find that the perfect-imperfect phase transition point lies on the critical line, separating molten and glassy regions, and coincides with the freezing transition at zero temperature. Therefore, while on the corresponding phase diagram the molten phase exists in both perfect and imperfect regions, the glassy phase is present only in the region with gaps.

The paper is organized as follows. First, we give an estimation of the transition value p_c by mapping of the problem to the so-called Dyck paths and estimating the fraction of "essential" arcs. Then we formulate the planar matching problem in terms of the matrix field theory proposed in Ref. [24] and discuss the self-consistent mean-field approximation. The estimations of the critical point, p_c, obtained
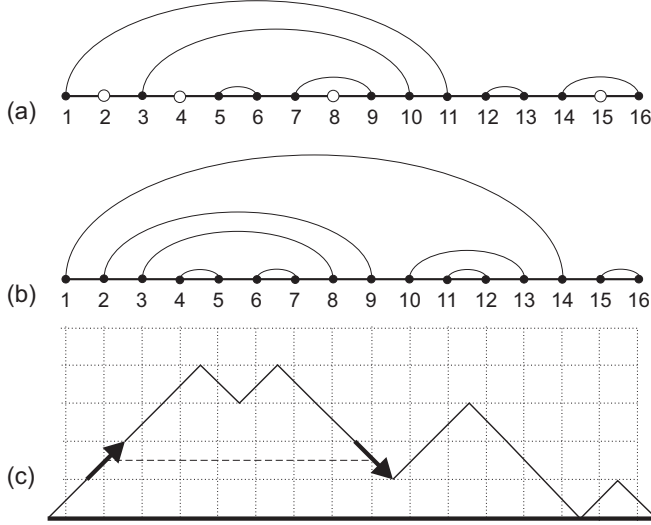
FIG. 1. Examples of imperfect (a) and perfect (b) planar matchings (open dots represent gaps); the gapless representation is shown by a Dyck path (c). The arc is given by an "up" and "down" steps at the same height, shown by arrows $\nearrow$ and $\searrow$. The part of the walk between arrows is a Dyck path itself.

analytically are compared to the values computed numerically via an exact dynamical programming algorithm. Finally, we characterize the fluctuation behavior of the free energy and discuss the relation between the perfect-imperfect transition and the molten-glass phase transition in random RNA with quenched primary sequence.

II. MAPPING ON BROWNIAN EXCURSIONS AND NAIVE MEAN FIELD

An intuitive idea about the calculation of the critical value p_c can be obtained by considering the one-to-one mapping between the L-point planar diagrams and the L-step Brownian excursions (BEs), known as Motzkin paths [25] (these paths are also called "height diagrams" in the context of applications to RNA folding [15]). Within this mapping, the gapless (perfect) planar configurations correspond to BEs with no horizontal steps, also known as Dyck paths; see Fig. 1(b) and 1(c). The total number of Dyck paths of even length L is given by a Catalan number

$$C_{L/2} = \frac{L!}{\left(\frac{L}{2}\right)!\left(\frac{L}{2}-1\right)!} \sim L^{-3/2} 2^L, \qquad (2)$$

where the asymptotic expression is valid for $L \gg 1$; $C_{L/2}$ represents the number of possible planar diagrams in the *fully connected* case, corresponding to $p = 1$ in our model.

For $0 < p < 1$, some planar diagrams in the fully connected ensemble are forbidden. This reduces the number of possible planar configurations, which becomes zero below a certain value p_c. A naive estimation of p_c can be easily obtained via the following mean-field-like argument. Since each arc in the diagram is present with a probability p, the probability that the whole configuration is allowed, is given by $p^{L/2}$. Assuming that planar diagrams in the fully connected ensemble are *statistically independent*, we get the probability

$\mathcal{P}$ to have at least one perfect planar matching configuration:

$$\mathcal{P} = 1 - (1 - p^{L/2})^{C_{L/2}} = 1 - \exp(-p^{L/2}C_{L/2}), \qquad (3)$$

where the last equality is valid for $L \to \infty$. In this limit the probability $\mathcal{P}$ is equal to one for $p > p_c$, and to zero for $p < p_c$. The perfect-imperfect naive mean-field threshold p_c is thus given by the condition

$$\lim_{L \to \infty} p_c[C_{L/2}]^{2/L} = 1, \qquad (4)$$

yielding $p_c = 1/4$. However, here we have neglected the statistical correlations between different configurations in the fully connected ensemble of planar configurations. Therefore, it provides only a lower bound to the true value of p_c. A careful account for such correlations leads to a natural generalization of the critical condition (4):

$$\lim_{L \to \infty} \xi(p_c)[C_{L/2}]^{2/L} = 1, \quad \xi(p_c) = 1/4, \qquad (5)$$

where $\xi(p)$ is some weight (due to correlations) to be determined.

III. COMBINATORICS OF "CORNER COUNTING"

An estimation of $\xi(p)$, and therefore of p_c, can be obtained by exploiting the combinatorial properties of Dyck paths. The consideration below provides an intuitive understanding of the statistical reasons beyond the shift of the transition probability from the mean-field value $p_c = 1/4$.

Our estimation is based on the following observation: the probabilities to find different arcs in a perfect matching structure crucially depend on the lengths of arcs. Consider a perfect matching of $L/2$ arcs connecting L points. In the limit $L \to \infty$ the *local statistics* of "up" and "down" steps in a corresponding Dyck path becomes independent on the global constraint for the random walk to be a Brownian excursion. Using the bijection between Dyck paths and arc diagrams, we see that the arc is drawn between ith and jth steps if and only if the ith step is $\nearrow$ ("up") and jth step is the first step $\searrow$ ("down") at the same height after i, as is depicted in the Fig. 1(c). Therefore, the probability to find an arc from i to j in a randomly chosen diagram can be formally written as a "correlation function":

$$P(i,j) = \frac{\langle \nearrow |\mathcal{D}_{i+1,j-1}| \searrow \rangle}{2^{j-i+1}}. \qquad (6)$$

In this expression the denominator represents the total number of possible sequences from ith to jth step; $\mathcal{D}_{i+1,j-1}$ is a Dyck path between $(i+1)$-th to $(j-1)$-th steps: this part of the walk should be a Dyck path itself to return to the same spatial coordinate for the *first time* at jth step. The number of such Dyck paths is given by the Catalan number $C_{(j-i-1)/2}$. Thus, $P(i,j)$ depends only on $k = j - i$ and equals

$$P(i,j) = \frac{C_{(k-1)/2}}{2^{k+1}}, \qquad (7)$$

and they are nonzero for odd k only: $P(i,i+1) = 1/4$, $P(i,i+3) = 1/16$, $P(i,i+5) = 1/32$, etc. The whole set of $P(i,j)$ sums to $\sum_{k=1}^{\infty} P(i,i+k) = 1/2$, which has a meaning of a probability that i is a starting (rather than ending) point of an arc.

Thus, the fraction of the shortest arcs of length $k = 1$, $P(i, i + 1) = 1/4$, represented by "up corners" $\wedge$ in a Dyck path, is exceptionally high. Indeed, in a typical fully connected diagram one half of the arcs ($L/4$ out of $L/2$) correspond to such corners. Moreover, while a fraction of long arcs chosen in each particular diagram is decaying at $L \to \infty$ (the number of possible long arcs is of order L^2, so the fraction of those chosen in each structure, is of order L^{-1}), the fraction of corners converges to $1/4$ (there are $L - 1$ possible corners, $L/4$ of them appear in a typical structure). Therefore, the values of quenched weights $A_{i,i+1}$ assigned to the short arcs in our model influence the existence of a perfect arc structure in a crucial way. Now we estimate how this exceptional role of the subdiagonal values $A_{i,i+1}$ influences p_c.

Assume that the typical arc structure is constructed as follows: (1) take $1/4$ corners (from $L - 1$ possible places) such that none of them touch each other, (2) select remaining $L/2 - L/4 = L/4$ arcs at random from ensemble of *any longer* arcs. Since the total number of longer arcs is of order of $L^2 \gg L/4$, we assume that the quenched disorder in the entries $A_{i,j}$ away from the subdiagonal can be safely ignored, and the contribution from the longer arcs into $\xi(p)$ remains as it is in the mean-field case (each arc is allowed with a probability p independently of others), thus

$$\xi^{L/2}(p) = \underbrace{p^{L/4}}_{\text{long arcs}} \underbrace{\mathcal{P}_{\wedge}(p)}_{\text{corners}}. \tag{8}$$

To determine the contribution of corners, $\mathcal{P}_{\wedge}(p)$, note that this value has a meaning of a probability to take $L/4$ corners at random (respecting the nontouching constraint) in a way that all of them belong to the set of pL allowed ones. Due to the nontouching constraint the problem of distributing corners can be mapped onto a problem of choosing $L/4$ objects (corners) out of $3L/4$ ones ($L/4$ corners and $L - 2 \times L/4 = L/2$ unmatched vertices; see Fig. 2). The number of corresponding partitions $\mathcal{Z}$ is

$$\mathcal{Z} = C_{3L/4}^{L/4} = \frac{\frac{3L}{4}!}{\frac{L}{4}!\frac{L}{2}!}. \tag{9}$$

In the Bernoulli model, only the fraction p of arc positions is fixed. Because of the nontouching constraint, it is natural

to assume that of $3L/4$ positions in the "point-and-stick" representation in Fig. 2(b) only $p(L - L/4) = 3pL/4$ are allowed on average (i.e., correspond to unity weights in the connectivity matrix A). Thus, the number of allowed partitions can be estimated as

$$\mathcal{Z}(p) = C_{3pL/4}^{L/4} = \frac{\frac{3pL}{4}!}{\frac{L}{4}!\left(\frac{3pL}{4} - \frac{L}{4}\right)!}. \tag{10}$$

Here $\mathcal{Z}(p)$ is the average number of possibilities to distribute shortest nontouching arcs at a given fraction p of allowed arcs, and

$$\mathcal{P}_{\wedge}(p) = \frac{\mathcal{Z}(p)}{\mathcal{Z}} \tag{11}$$

is a probability, given p, to pick an allowed set of short arcs at random. Substituting (11) into (8) one gets in the limit $L \to \infty$ the following result for $\xi(p)$:

$$\ln \xi(p) = \frac{1}{2} \ln p + \frac{3p}{2} \ln \frac{3p}{2} - \frac{3p - 1}{2} \ln \frac{3p - 1}{2} - \frac{3}{2} \ln \frac{3}{2}. \tag{12}$$

Being substituted into (5), Eq. (12) gives an estimate for the transition point $p_c = 0.35$. We see therefore that the transition point shifts significantly from its mean-field value due to the special role of a quenched disorder in the subdiagonal entries $A_{i,i+1}$ of the connectivity matrix A.

IV. SELF-CONSISTENT FIELD THEORY FOR PLANAR ARC COUNTING

A different way to attack the planar matching problem consists in using the matrix model approach proposed in Ref. [24] and the $1/N$ expansion, a standard technique in quantum field theory. For the set of L vertices, associate to a vertex i an Hermitian matrix $(\phi_i)_{N \times N}$. The L-point generating function Z_L can be written as follows:

$$Z_L(N; A) = \frac{\int d\phi_1 \cdots d\phi_L e^{-H_0} \frac{1}{N} \text{tr}(\phi_1 \cdots \phi_L)}{\int d\phi_1 \cdots d\phi_L e^{-H_0}}$$
$$\equiv \langle \phi_1 \cdots \phi_L \rangle_{H_0}, \tag{13}$$

where

$$H_0 = \frac{N}{2} \sum_{i,j} (A^{-1})_{ij} \text{tr}(\phi_i \phi_j). \tag{14}$$

Since $\text{tr}(\phi_i \phi_j) = \sum_{a,b} \phi_{ab}^i \phi_{ba}^j = \sum_{a,b,c,d} \delta_{ad} \delta_{bc} \phi_{ab}^i \phi_{cd}^j$, every propagator enters with a $1/N$ factor, while every loop gives a N factor. Due to the Wick theorem, one has

$$\langle \phi_1 \cdots \phi_L \rangle_{H_0} = \left\langle \sum_{\text{pairs}} \prod_{k,k'} \phi_k \phi_{k'} \right\rangle_{H_0}, \tag{15}$$

where each nonplanar configuration comes with a factor $1/N^2$ to some power and therefore vanishes in the $N \to \infty$ limit. Thus, the generating function $Z_L(N; A)$ counts in the limit $N \to \infty$ the number of planar diagrams with exactly $L/2$ arcs (on genus $g = 0$ surface) compatible with a specific realization of the disorder defined by the matrix A. In the absence of any disorder, one can set $A_{ij} \equiv \alpha$ for any (i, j), where α is some constant (it corresponds to the $p = 1$ limiting case). In this case

shortest arcs (L/4 of them)

(a)

⇕

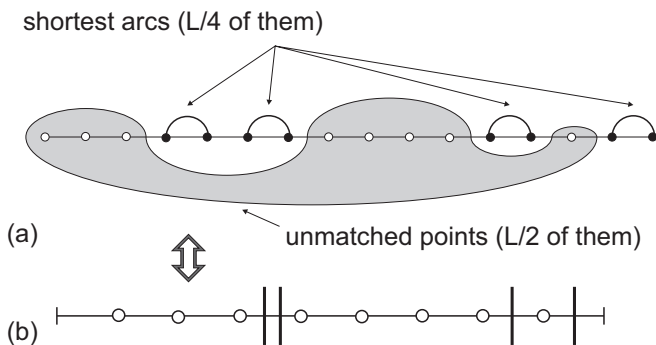**(b)**

unmatched points (L/2 of them)

FIG. 2. Computation of $\mathcal{Z}$ and $\mathcal{Z}(p)$: (a) Selection of $L/4$ nontouching arcs on the set of L points ($L/2$ open dots remain unmatched); (b) the same problem reformulated as a partitioning of vertical segments (arcs) between open dots (unmatched points). A certain number of partitions are forbidden by the matrix of contacts A.

the multidimensional integral (13) can be reduced by a series of Hubbard-Stratonovich transformations to a one-dimensional integral involving the spectral density of a Gaussian matrix, which is a well-known result of random matrix theory (RMT). We will refer to this realization of A as to the fully connected case. If we set $\alpha = 1$, we get [24]

$$\lim_{N \to \infty} Z_L(N; A) = C_{L/2}, \tag{16}$$

where $C_{L/2}$ is the Catalan number (2), as it should be. However, for a generic disordered matrix A, the calculations are intractable. Still, we show below that by averaging over the matrix distribution (1) and by applying the self-consistency arguments, we are able to treat the partially connected system with $0 < p < 1$ as an effective fully connected system with α different from one, thus obtaining a correction to the naive mean-field result (4).

According to the consideration above, the function $\xi(p)$ defined by Eq. (5) can be calculated within the matrix approach by averaging $Z_L(N; A)$ over the distribution (1). To this end, we use the standard Hubbard-Stratonovich transform and integrate over A with the weight (1):

$$\int dA\, P(A)\, Z_L(N; A) = C \int \prod_{k=1}^{L} d\phi_k \frac{1}{N} \mathrm{tr}(\phi_1 \ldots \phi_L)$$

$$\times \int \prod_{m=1}^{L} dh_m e^{iN \sum_i \mathrm{tr}(h_i \phi_i)} e^S, \tag{17}$$

where C is a constant, $S = S_0 + V$, and

$$S_0 = -\frac{pN}{2} \sum_{ij} \mathrm{tr}(h_i h_j), \tag{18}$$

$$V = \frac{p(1-p)N^2}{8} \sum_{ij} [\mathrm{tr}(h_i h_j)]^2 - \frac{p(1-p)(1-2p)N^3}{48}$$

$$\times \sum_{ij} [\mathrm{tr}(h_i h_j)]^3 + \cdots. \tag{19}$$

Up to this point, no approximation has been made. The S_0 term (18) corresponds to a fully connected matrix with an additional factor p behind. If this term is the only present, then, performing the inverse Hubbard-Stratonovich transformation and returning to the functional of the type (13), we get $\xi(p) = p$, recovering the value $p_c = 1/4$ given by the critical condition (4).

The correction to p_c due to the rest of the series V (19) can be estimated as follows. The series given by the action S can be thought of as a Gaussian theory with the interaction V. Since V contains an infinite number of terms, it is impossible to treat it perturbatively. Still, we can use a self-consistent nonperturbative approach reminiscent of Feynman's variational principle [26] in the field theory: as all the fields $\{h_i\}_{i=1,\ldots,L}$ in Eq. (19) are equivalent, we assume that the average $\langle N \mathrm{tr}(h_i h_j) \rangle_{S_0} \equiv U$ is independent on (i, j). Within the adopted mean-field approximation, the replacement

$e^S = e^{S_0} e^{\langle V \rangle}$ is supposed, where

$$\langle V \rangle = \frac{p(1-p)N}{8} U \sum_{ij} \mathrm{tr}(h_i h_j) - \frac{p(1-p)(1-2p)N}{48} U^2$$

$$\times \sum_{ij} \mathrm{tr}(h_i h_j) + \cdots. \tag{20}$$

Resumming the series (20), we obtain the following self-consistent equation for the "propagator" U:

$$\frac{1}{U} = -\frac{2}{U} \log \left[1 - p + p \exp \left(-\frac{U}{2} \right) \right]. \tag{21}$$

The equation (21) yields $U = -2 \log [1 - \frac{1 - 1/\sqrt{e}}{p}]$. Hence, finally, we can write

$$S = -\frac{\xi(p)N}{2} \sum_{ij} \mathrm{tr}(h_i h_j), \tag{22}$$

where

$$\xi(p) = \left(-2 \log \left[1 - \frac{1 - 1/\sqrt{e}}{p} \right] \right)^{-1}. \tag{23}$$

Substituting (23) into (5), we get an estimation for the critical value $p_c^* = 0.455$. Although the self-consistent approximation (20) seems to be rather crude (the numerical estimation of p_c for large matrices is $p_c \approx 0.379$), it leads to the correct direction of the shift of p_c from the naive mean-field value $p_c = 0.25$ (4). It would be interesting to understand how to treat the interaction term V (19) more properly.

V. PERFECT-IMPERFECT PHASE TRANSITION

The combinatorial problem of planar matching, introduced above, is strongly related to the problem of optimal folding of random RNAs. A real RNA represents a single-stranded polymer, composed of four types of nucleotides: A, C, G, and U. Under normal conditions, the RNA molecule folds onto itself and forms double-helical structures of stacked base pairs, known as the *secondary structure* of the RNA, favoring the stable Watson-Crick pairs A-U and G-C. The simplest theories describe the statistics of the *random* RNA secondary structures, incorporating the most important features: saturation of base pairings, exclusion of the so-called pseudoknots, which are known to be very rare in real RNA [11], and, sometimes, the condition of finite flexibility of the molecule, requiring a minimal length of a loop [22,23]. The exclusion of the pseudoknots means that the base pairings can be represented by one-dimensional planar diagrams, depicted in Fig. 1. This topological constraint allows us to calculate the partition function of the RNA, using an exact dynamical programming algorithm [9,11]. The recursion relation for the partition function, $Z_{i,i+k}$, of the part between monomers i and $i + k$, reads

$$Z_{i,i+k} = Z_{i+1,i+k} + \sum_{s=i+1}^{i+k} \beta_{i,s} Z_{i+1,s-1} Z_{s+1,i+k}, \tag{24}$$

where $\beta_{i,j} = e^{-A_{i,j}/T}$ are statistical weights of bonds ($1 \leqslant i < j \leqslant n$); $A_{i,j} = 1$ if i and j match each other, and $A_{i,j} = 0$ otherwise. In the zero-temperature limit, Eq. (24) is reduced

to the dynamical programming algorithm for the ground state free energy [27]:

$$
\begin{aligned}
F_{i,i+k} &= \lim_{T \to 0} T \ln Z_{i,i+k} \\
&= \max_{s=i+1,\ldots,i+k} \{F_{i+1,i+k}, \varepsilon_{i,s} + F_{i+1,s-1} + F_{s+1,i+k}\}.
\end{aligned}
\tag{25}
$$

Since the free energy, F, of the whole chain is proportional to the number of nucleotides involved in the planar bond formation, the combinatorial problem of planar matching can be regarded as a $T = 0$ optimization problem for the free energy of the RNA molecule with a given matrix of contacts, A. Therefore, the exact dynamical programming algorithm (25) allows us to detect the phase transition by considering the fraction $f_L(p) = 2F/L$ of links, involved in planar binding, for different densities of contacts p in the limit $L \to \infty$: one expects $f_\infty(p) = 1$ for $p > p_c$, and $f_\infty(p) < 1$ for $p < p_c$.

Thus, looking for the fraction, $\eta_L(p)$, of sequences, which allow perfect matchings, in the whole ensemble of random sequences, one has $\eta_\infty(p) = 1$ for $p > p_c$, and $\eta_\infty(p) = 0$ for $p < p_c$. The corresponding dependencies are shown in Fig. 3(a) for different polymer lengths, $L = 500, 1000, 2000$. As $L \to \infty$, the function $\eta_L(p)$ tends to a step function. Two different phases are observed: for $p > p_c$ one has a gapless perfect matching with all nucleotides involved in planar binding, while for $p < p_c$ there is always a finite fraction of gaps in the best possible matching.

The scaling analysis permits to determine the phase transition point as $p_c \approx 0.379$ (compare to the predictions of the "corner counting" and of the self-consistent field theory). Fig. 3(b) shows that curves with different L collapse, demonstrating the scaling behavior $\eta[(p - p_c)/L^\nu]$ with the transition width $L^{-\nu}$, where $\nu = 0.5$. The convergence of the function f_L to the limiting value $f_\infty(p)$ (see Fig. 4) in the perfect and imperfect phases has, respectively, exponential and power-law tails:

$$
\begin{aligned}
f_\infty(p) - f_L(p) &\sim e^{-L/\ell(p)} \quad \text{for } p > p_c \\
f_\infty(p) - f_L(p) &\sim L^{-\alpha(p)} \quad \text{for } p < p_c.
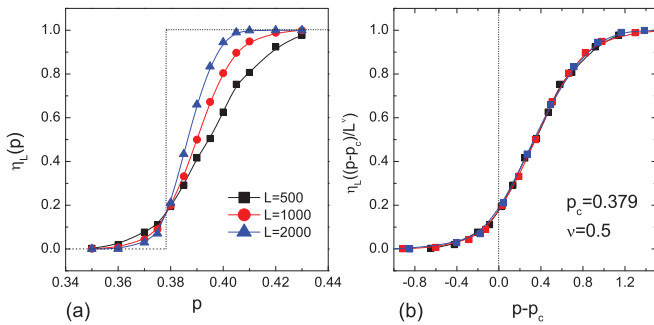\end{aligned}
\tag{26}
$$

FIG. 3. (Color online) (a) The fraction of perfect matchings $\eta_L(p)$ as a function of the density p of ones in the contact matrix A for chain lengths $L = 500, 1000, 2000$, averaged over 10 000 instances. The dashed line corresponds to the thermodynamic limit $L \to \infty$, yielding the critical value $p_c = 0.379$. (b) The scaling analysis of curves, corresponding to different chain lengths L. The fitting procedure gives the exponent of the transition width $\nu \approx 0.5$.
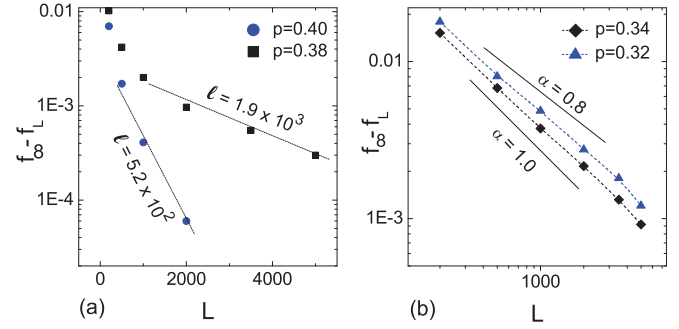

FIG. 4. (Color online) Convergence of fraction of links, involved in planar binding, f_L to the limiting value f_∞ in two regimes, $p > p_c$ and $p < p_c$. (a) In the perfect phase, the exponential convergence is demonstrated for $p = 0.38$ and $p = 0.4$ in the semilogarithmic scale. The screening length $\ell(p)$ diverges as p approaches the critical value p_c. (b) In the imperfect phase, the power-law behavior is shown for $p = 0.32$ and $p = 0.34$ in the log-log scale. The exponent $\alpha(p)$ as a function of p takes values between 0.8 and 1. The data points are averaged over 1000 instances.

In the perfect matching phase the screening length $\ell(p)$ diverges at the point $p = p_c$ [two examples for $p = 0.38$ and $p = 0.4$ are shown in the Fig. 4(a) in the semilogarithmic scale], while for the imperfect matching the finite-size scaling analysis demonstrates the power-law behavior with the exponents $0.8 \leqslant \alpha(p) \leqslant 1$ [see Fig. 4(b) for two examples, $p = 0.32$ and $p = 0.34$ in the log-log plot]. Note that the exponential scaling for $p > p_c$ may not be universal (being model dependent) and is likely to be a feature of the Bernoulli model (1), while the power-law behavior for $p < p_c$ appears in other models, e.g., for integer-valued "alphabet" [21].

VI. MOLTEN-GLASS PHASE TRANSITION

The investigation of thermodynamic properties of RNA secondary structures has been addressed in a number of papers [12,13,15,18,19,22,23]. Many of them provided numerical and analytical evidence for existence of a low-temperature glassy phase. In Ref. [15] it was shown that in the high-temperature phase the system remains in the molten phase, characterized by a homopolymer-like behavior. In the molten phase the disorder is irrelevant, and the binding matrix elements A_{ij} can be replaced by some effective value α. Carrying out the two-replica calculation, the authors were able to prove that the system exhibits a phase transition from a high-temperature regime, in which the replicas are independent, to a low-temperature phase, in which the disorder is relevant and replicas are strongly coupled. The authors numerically characterized the transition to a glassy phase by imposing a pinch between two bases and measuring the corresponding energy cost.

Several other works [22,23] used an alternative so-called ε-coupling method, to investigate the nature and the scaling laws of the glassy phase, observing the effect of typical excitations imposed by a bulk perturbation. The authors argued that for the models with nondegenerate ground states, the low-temperature phase is not marginal, but is governed by a scaling exponent, close to $\theta = 1/3$. The explicit numerical studies of the specific

heat demonstrate that molten-glass transition is only a fourth order phase transition [13].

Regardless of particular details of models considered in all these works, it is clear that the existence of the glassy phase is possible only in a sufficiently disordered and frustrated system. Besides the planarity constraint, shared by all simple models of random RNA, the Bernoulli model is described by a unique disorder parameter, p, that controls the density of allowed contacts. In this model the appearance of the glassy phase is impossible above a certain threshold, p^*. Indeed, it is well known that for $p = 1/2$ (corresponding to an effective alphabet $c = 1/p = 2$), there is no transition to the glassy phase at all, and the system remains always in the molten phase [13,15]. Below we present arguments supporting the hypothesis that p^* is equal to the critical value of the perfect-imperfect matching transition, p_c, discussed above.

To identify the dependence of the molten-glass transition temperature on the effective alphabet (defined as $c = 1/p$), we follow the procedure suggested in Ref. [15]. In the high-temperature regime the disorder is irrelevant (this corresponds to a homopolymer-like behavior in polymer language) and one can put $A_{ij} = \alpha$. In this regime the free energy of the chain of length L scales linearly with L, up to a logarithmic correction, which is just the logarithm of the power-law multiplier in the Catalan number (2) enumerating all possible structures: $F(L,T) = f(T)L - (3T/2)\ln L$, where $f(T)$ is some (nonuniversal) function of the temperature. In particular, the energy cost of imposing a bond connecting two monomers at distance $L/2$ from each other equals in the high-temperature regime

$$\Delta F(L,T) = F(L,T) - 2F(L/2,T) = \frac{3}{2}T\ln\frac{L}{4}. \quad (27)$$

The violation of this behavior indicates [15] the appearance of the glassy phase. This fact can be used to detect the transition temperature in the Bernoulli model. Namely, we use the following fit for $\Delta F(L,T)$ [where $F(L,T)$ is to be determined via recursion relations (25)],

$$\Delta F(L,T) = a(T)\ln L + b(L), \quad (28)$$

and plot the T dependence of $a(T)$; see Fig. 5. We interpret the deviation of the $a(T)$ from the high-temperature value $3T/2$ as appearance of the glass transition. Note that the logarithmic fit (28) for the free energy does not give a correct asymptotics at low temperatures (indeed, the true asymptotics is known to include power-law and logarithm-squared terms [19]).

As follows from Fig. 5, the expected behavior (27) is indeed observed at high temperatures and is violated at a certain temperature T_c. Following Ref. [15] we identify this regime change with the molten-glass transition. We see that with the increase of p, the critical temperature T_c shifts to lower values, approaching zero for some $0.35 < p^* < 0.5$. At low temperatures, the numerical computations become very time consuming, leading to the loss of precision in the vicinity of p^*. However, it seems that the hypothesis $p^* = p_c$ still holds: the sequences corresponding to $p > p_c$ remain in the molten phase, the pinching free energy (28) has the same dependence even for very low temperatures.
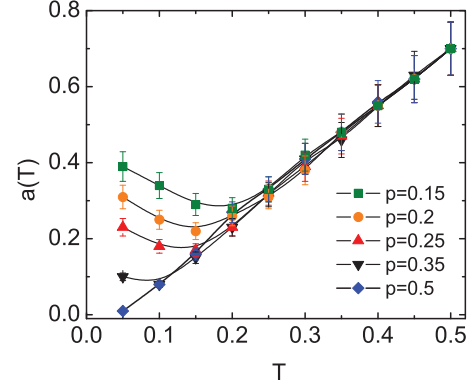

FIG. 5. (Color online) The dependence of the coefficient $a(T)$ in (28) on the temperature for $p = 0.15$, 0.2, 0.25, 0.35, 0.5. For $p > p^*$ ($0.35 < p^* < 0.5$), the coefficient $a(T)$ seems to follow the $a(T) = 3T/2$ law, typical for the molten phase, up to very low temperatures. For $p < p^*$, the $a(T)$ dependence deviates from the high-temperature behavior at some temperature, which we identify as a critical temperature of transition to the glassy phase. The data points are averaged over 10 000 samples.

VII. DISCUSSION: MATCHING VS FREEZING

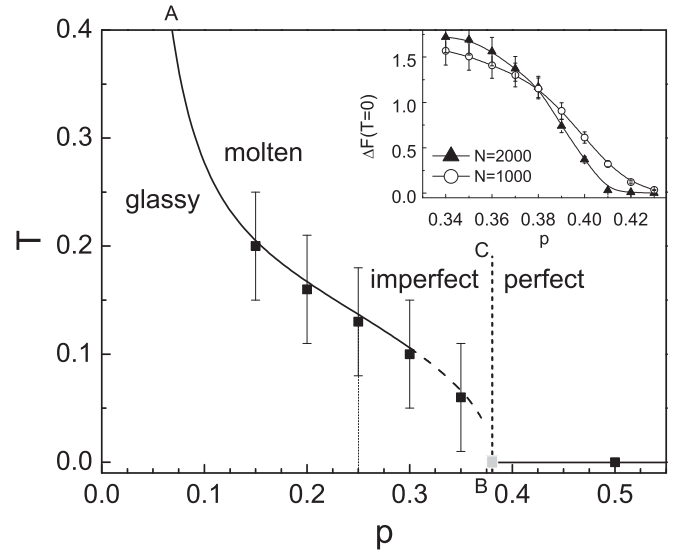The results presented in this work suggests the generic phase diagram shown in Fig. 6 for the Bernoulli model of

FIG. 6. The phase diagram of the Bernoulli model on the (T, p) plane. The data points correspond to the critical temperature T_c of the molten-glass transition for different values $p = 0.15$, 0.2, 0.25, 0.3, 0.35, 0.5. A four-letter alphabet ($p = 0.25$) is highlighted by a thin dashed line. The critical curve (A–B) separates glassy and molten phases. We conjecture that at zero temperature, the endpoint B, giving p^*, coincides with the critical point p_c for the perfect-imperfect transition. The thick dashed line (B–C) separates the perfect and imperfect matching cases. The glassy phase lies entirely inside the region, characterized by gaps. Inset: Evidence for the conjecture $p^* = p_c$. Study of the pinching free energy $\Delta F(L,T)$ at zero temperature. In the limit of large L, the glassy phase is absent for $p > p^*$, characterized by $\Delta F(\infty,0) = 0$. The point p^* can be identified as a crossing point for different $\Delta F(L,0)$ curves, presented here for $L = 1000$ and $L = 2000$, and its value is found to be very close to $p_c = 0.379$. The data points are averaged over 1000 samples.

random RNA chains. The perfect-imperfect transition at zero temperature separates two matching regions: with and without gaps. Analytically, we proved the existence of the transition from the perfect matching region to the imperfect one, and we provided estimates for the values of the transition point, p_c. Using the exact dynamical programming algorithm (25), we found this critical value to be $p_c \approx 0.379$, highlighted by a thick dashed line (B–C) in Fig. 6. The previous studies have been mostly concentrated on the description of the finite-temperature molten-glass transition for a sufficiently frustrated model with a fixed alphabet (a fixed p in the Bernoulli model). An example of such a phase transition point is marked by a thin dashed line in the Fig. 6 and corresponds to an intensively studied case of the four-letter alphabet ($p = 0.25$). The ensemble of critical points for different values of p gives a critical curve (A–B) in the (T, p) plane.

The computational cost increases drastically for temperatures close to zero (and, hence, in the vicinity of p_c), and the recursive relations (24) are no more applicable. However, we can still try to carry out the analysis of the pinching free energy $\Delta F(L, T)$ at zero temperature, using the exact dynamical programming algorithm (25). Indeed, the glassy phase does not exist if $\Delta F(\infty, 0) = 0$. This happens for $p > p^*$, where p^* is defined as the density of constrains, for which the critical temperature is zero: $T_c(p^*) = 0$. The corresponding plot is shown in the inset of Fig. 6. According to (26), the pinching free energy (27) decreases with growth of L in the imperfect matching phase, while increases (with growth of L) in the perfect matching regime. Hence, the value of p^* in the large L limit can be identified as a crossing point of $\Delta F(L, 0)$ curves for different L. The crossing point for $L = 1000$ and $L = 2000$ is indeed found to be very close

to the value $p_c = 0.379$, strongly supporting the hypothesis $p^* = p_c$. The aforementioned results indicate that the critical curve $T_c(p)$ crosses zero at the critical value p_c. Hence, the perfect-imperfect transition point seems to lie at the critical line, separating molten and glassy regions, and coincides with its limiting $T = 0$ value. We see that although the glassy phase exists only in the region where the gaps are present, the molten phase lies in both, perfect and imperfect, matching regions.

Because of the one-parameter dependence, the Bernoulli model is probably the simplest model for modeling the secondary structure of the RNA, that captures the essential physical properties of the process. Being applied to the studies of the thermodynamic properties of random RNAs, the problem introduced in this paper provides some enlightenment on the nature of molten-glass transition at zero temperature. Starting from the Bernoulli model, one could directly generalize our approach to investigate more sophisticated and realistic models of the RNA secondary structure, for example, by introducing the minimal allowed hairpin length [13,15,23], taking into account the pseudoknots [24] and different binding probabilities [22,24].

[1] E. Friedgut, J. Amer. Math. Soc. **12**, 1017 (1999).

[2] E. Brézin, C. Itzykson, G. Parisi, and J. B. Zuber, Commun. Math. Phys. **59**, 35 (1978).

[3] A. A. Abrikosov and L. P. Gorkov, *Methods of Quantum Field Theory in Statistical Physics* (Courier Dover Publications, New York, 1975).

[4] R. Saito, J. Phys. Soc. Jpn. **59**, 482 (1990).

[5] M. L. Mehta, *Random Matrices* (Academic Press, New York, 2004).

[6] G. Rodgers and A. Bray, Phys. Rev. B **37**, 3557 (1988).

[7] A. D. Mirlin and Y. V. Fyodorov, J. Phys. A **24**, 2273 (1991).

[8] G. Semerjian and L. F. Cugliandolo, J. Phys. A **35**, 4837 (2002).

[9] P. G. de Gennes, Biopolymers **6**, 715 (1968).

[10] I. Ya. Erukhimovich, Vysokomol. Soyed. B **20**, 10 (1978).

[11] R. Nussinov and A. B. Jacobsont, Proc. Natl. Acad. Sci, U.S.A. **77**, 6309 (1980).

[12] R. Bundschuh and T. Hwa, Phys. Rev. Lett. **83**, 1479 (1999).

[13] A. Pagnani, G. Parisi, and F. Ricci-Tersenghi, Phys. Rev. Lett. **84**, 2026 (2000).

[14] A. Montanari and M. Mezard, Phys. Rev. Lett. **86**, 2178 (2001).

[15] R. Bundschuh and T. Hwa, Phys. Rev. E **65**, 031903 (2002).

[16] H. Orland and A. Zee, Nucl. Phys. B **620**, 456 (2002).

[17] M. Müller, Phys. Rev. E **67**, 021914 (2003).

[18] M. Lässig and K. J. Wiese, Phys. Rev. Lett. **96**, 228101 (2006).

[19] S. Hui and L.-H. Tang, Eur. Phys. J. B **53**, 77 (2006).

[20] M. V. Tamm and S. K. Nechaev, Phys. Rev. E **75**, 031904 (2007).

[21] O. V. Valba, M. V. Tamm, and S. K. Nechaev, Phys. Rev. Lett. **109**, 018102 (2012).

[22] E. Marinari, A. Pagnani, and F. Ricci-Tersenghi, Phys. Rev. E **65**, 041919 (2002).

[23] F. Krzakala, M. Mézard, and M. Müller, Europhys. Lett. **57**, 752 (2002).

[24] G. Vernizzi, H. Orland, and A. Zee, Phys. Rev. Lett. **94**, 168103 (2005).

[25] S. K. Lando, *Lectures on Generating Functions* (American Mathematical Society, Providence, RI, 2003).

[26] R. Feynman, Phys. Rev. **97**, 660 (1955).

[27] S. K. Nechaev, M. V. Tamm and O. V. Valba, J. Phys. A **44**, 195001 (2011).