Bethe Ansatz in the Bernoulli matching model of random sequence alignment

Satya N. Majumdar,[1] Kirone Mallick,[2] and Sergei Nechaev[1,*]
[1]*Laboratoire de Physique Théorique et Modèles Statistiques, Université de Paris-Sud, CNRS UMR 8626, 91405 Orsay Cedex, France*
[2]*Service de Physique Théorique, Saclay, 91191 Gif-sur-Yvette cedex, France*
(Received 3 October 2007; published 11 January 2008)

For the Bernoulli matching model of the sequence alignment problem we apply the Bethe Ansatz technique via an exact mapping to the five-vertex model on a square lattice. Considering the terracelike representation of the sequence alignment problem, we reproduce by the Bethe Ansatz the results for the averaged length of the longest common subsequence in the Bernoulli approximation. In addition, we compute the average number of nucleation centers of the terraces.

DOI: 10.1103/PhysRevE.77.011110 PACS number(s): 02.50.−r, 87.10.−e, 87.15.Cc, 05.40.−a

I. INTRODUCTION: BERNOULLI MATCHING MODEL OF SEQUENCE ALIGNMENT

The goal of a sequence alignment problem is to find similarities in patterns in different sequences. Sequence alignment is one of the most useful quantitative methods of evolutionary molecular biology [1–3]. A classic alignment problem deals with the search of the longest common subsequence (LCS) in two random sequences. Finding analytically the statistics of LCS of a pair of sequences randomly drawn from the alphabet of c letters is a challenging problem in computational evolutionary biology. The exact asymptotic results for the distribution of LCS have been derived recently in [4] in a simpler, yet nontrivial, variant called the Bernoulli matching (BM) model (see details below). It has been shown in [4] via a sequence of mappings that in the BM model, for all c, the distribution of the asymptotic length of the LCS, suitably scaled, is identical to the Tracy-Widom distribution of the largest eigenvalue of a random matrix whose entries are drawn from a Gaussian unitary ensemble [5,6].

The problem of finding the longest common subsequence in a pair of sequences drawn from the alphabet of c letters is explicitly formulated as follows. Consider two sequences $\alpha=\{\alpha_1,\alpha_2,\ldots,\alpha_i\}$ (of length i) and $\beta=\{\beta_1,\beta_2,\ldots,\beta_j\}$ (of length j). For example, let α and β be two random strings of $c=4$ base pairs A,C,G,T of a DNA molecule, e.g., $\alpha=\{A,C,G,C,T,A,C\}$ with $i=6$ and $\beta=\{C,T,G,A,C\}$ with $j=5$. Any subsequence of α (or β) is an ordered sublist of α (or β), i.e., subsequences which need not be consecutive. For example, $\{C,G,T,C\}$ is a subsequence of α, but $\{T,G,C\}$ is not. A common subsequence of two sequences α and β is a subsequence of both of them. For example, the subsequence $\{C,G,A,C\}$ is a common subsequence of both α and β. There are many possible common subsequences of a pair of sequences. The aim of the LCS problem is to find the longest of them. This problem and its variants have been widely studied in biology [7–11], computer science [2,12–14], probability theory [15–21], and more recently in statistical physics [22,23]. A particularly important application of the LCS problem is to quantify the closeness between two DNA sequences. In evolutionary biology, the genes responsible for building specific proteins evolve with time and by finding the LCS of the same gene in different species, one can learn what has been conserved in time. Also, when a new DNA molecule is sequenced *in vitro*, it is important to know whether it is really new or it already exists. This is achieved quantitatively by measuring the LCS of the new molecule with another existing molecule already in the database.

Computationally, the easiest way to determine the length $L_{i,j}$ of the LCS of two arbitrary sequences of lengths i and j [in polynomial time $\sim O(ij)$] with no cost of gaps can be achieved using the simple recursive algorithm [2,23]

$$L_{i,j}=\max[L_{i-1,j},L_{i,j-1},L_{i-1,j-1}+\eta_{i,j}], \tag{1}$$

subject to the initial conditions $L_{i,0}=L_{0,j}=L_{0,0}=0$, where the variable $\eta_{i,j}$ is

$$\eta_{i,j}=\begin{cases}1 & \text{if characters at the positions } i \text{ (in } \alpha) \\ & \text{and } j \text{ (in } \beta) \text{ match each other,} \\ 0 & \text{otherwise,}\end{cases} \tag{2}$$

In Fig. 1(a) the matrix of the variables $\eta_{i,j}$ is shown for a particular pair of sequences $\alpha=\{A,C,G,C,T,A,C\}$ and $\beta=\{C,T,G,A,C\}$ discussed above. Any common subsequence can be represented by a *directed* path connecting sequentially $\{\eta_{i_1,j_1},\eta_{i_2,j_2},\ldots,\eta_{i_s,j_s},\ldots,\eta_{i_n,j_n}\}$, where $\eta_{i_1,j_1}=\eta_{i_2,j_2}=\cdots=\eta_{i_s,j_s}=\cdots=\eta_{i_n,j_n}=1$ and $i_{s+1}>i_s$, $j_{s+1}>j_s$ for all $1\le s\le n$. Two particular realizations of common subsequences, namely $\{C,G,A,C\}$ and $\{A,C\}$, are shown in Fig. 1(a) by two broken lines connecting 1 (the first one is the one of LCSs). In Fig. 1(b) we show the table of all $L_{i,j}$ ($0\le i\le 6$, $0\le j\le 5$) corresponding to the matrix $\eta_{i,j}$ (the first line and the first column are the boundary conditions $L_{0,j}=L_{j,0}=0$). Let us note the terracelike structure of Fig. 1(b): the numbers from 0 to 4 could be viewed as different heights of the terraces. We shall later address to this representation.

For a pair of fixed sequences of lengths i and j, respectively, the length $L_{i,j}$ of their LCS is just a number. However, in the statistical version of the LCS problem one compares two random sequences drawn from the alphabet of c letters and hence the length $L_{i,j}$ is a random variable. The statistics of $L_{i,j}$ has been intensively studied during the past three de-

*Current address: P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991, Moscow, Russia.

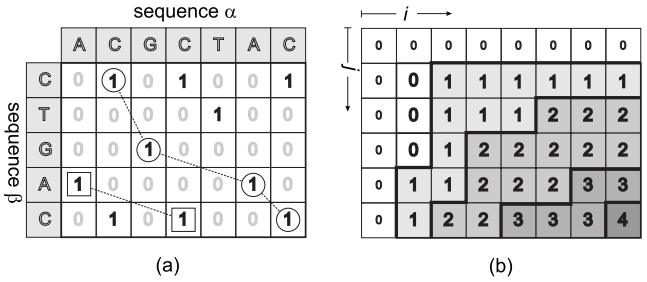

FIG. 1. (a) Matrix of variables $\eta_{i,j}$; (b) table of lengths $L_{i,j}$ of all LCSs corresponding to (a).

cades [15–19]. For equally long sequences ($i=j=n$), it has been proved that $\langle L_{n,n}\rangle \approx \gamma_c n$ for $n \gg 1$, where the averaging is performed over all uniformly distributed random sequences. The constant γ_c is known as the Chvátal-Sankoff constant which, to date, remains undetermined though there exists several bounds [16,18,19], a conjecture due to Steele [17] that $\gamma_c = 2/(1+\sqrt{c})$, and a recent proof [20] that $\gamma_c \to 2/\sqrt{c}$ as $c \to \infty$. Unfortunately, no exact results are available for the finite size corrections to the leading behavior of the average $\langle L_{n,n}\rangle$, for the variance, and also for the full probability distribution of $L_{n,n}$. Thus, despite tremendous analytical and numerical efforts, exact solution of the random LCS problem is far from being completely resolved. One feature that makes this problem particularly complicated is that the variables $\eta_{i,j}$ that are defined in (2) are not mutually independent but are correlated. To see that consider the simple example—matching of two strings $\alpha=AB$ and $\beta=AA$. One has by definition $\eta_{1,1}=\eta_{1,2}=1$ and $\eta_{2,1}=0$. The knowledge of these three variables is sufficient to predict that the last two letters do not match each other, i.e., $\eta_{2,2}=0$. Thus, $\eta_{2,2}$ cannot take its value independently of $\eta_{1,1}, \eta_{1,2}, \eta_{2,1}$. Note however that for two random sequences drawn from the alphabet of c letters, the correlations between the $\eta_{i,j}$ variables vanish in the $c \to \infty$ limit.

A first natural question is whether the problem is solvable in the absence of correlations between the $\eta_{i,j}$'s? This question leads to the Bernoulli matching (BM) model which is a simpler variant of the original LCS problem where one ignores the correlations between $\eta_{i,j}$'s for all c [23]. The length $L_{i,j}^{\mathrm{BM}}$ of the BM model satisfies the same recursion relation as in Eq. (1) except that $\eta_{i,j}$'s are now independent and each $\eta_{i,j}$ is drawn from the bimodal distribution,

$$p(\eta) = \frac{1}{c}\delta_{\eta,1} + \left(1 - \frac{1}{c}\right)\delta_{\eta,0} = \begin{cases} \dfrac{1}{c} & \text{for } \eta = 1, \\ 1 - \dfrac{1}{c} & \text{for } \eta = 0. \end{cases} \tag{3}$$

This approximation is expected to be exact only in the $c \to \infty$ limit. Nevertheless, for finite c, the results on the BM model can serve as a useful benchmark for the original LCS model to decide if indeed the correlations between $\eta_{i,j}$'s are important or not. Progress has been made for the BM model which, though simpler than the original LCS model, is still nontrivial. The average matching length $\langle L_{n,n}^{\mathrm{BM}}\rangle$ in the BM model, for large sequence lengths, n, was first computed by Seppäläinen [21] using probabilistic method and it was shown that $\langle L_{n,n}^{\mathrm{BM}}\rangle \approx \gamma_c^{\mathrm{BM}} n$ for $n \gg 1$ where $\gamma_c^{\mathrm{BM}} = 2/(1+\sqrt{c})$, same as the Steele's conjectured value, γ_c, for the original LCS model. Later the same result was rederived [23] in the physics literature using the cavity method of the spin-glass physics. Recently, in [4], we derived the asymptotic limit law for the distribution of the random variable $L_{n,n}^{\mathrm{BM}}$ and showed that for large n we have

$$L_{n,n}^{\mathrm{BM}} \to \gamma_c^{\mathrm{BM}} n + f(c) n^{1/3} \chi, \tag{4}$$

where $\gamma_c^{\mathrm{BM}} = 2/(1+\sqrt{c})$ and χ is a random variable with a n-independent distribution, $\mathrm{Prob}(\chi \le x) = F_{\mathrm{TW}}(x)$, which is the well-studied Tracy-Widom distribution for the largest eigenvalue of a random matrix with entries drawn from a Gaussian unitary ensemble [5]. For a detailed form of the function $F_{\mathrm{TW}}(x)$, see [5]. We also have shown that for all c,

$$f(c) = \frac{c^{1/6}(\sqrt{c}-1)^{1/3}}{\sqrt{c}+1}. \tag{5}$$

This allowed us to calculate in [4] the average length $L_{n,n}$ including the subleading finite size correction term, as well as the variance of $L_{n,n}^{\mathrm{BM}}$ for large n,

$$\langle L_{n,n}^{\mathrm{BM}}\rangle \approx \gamma_c^{\mathrm{BM}} n + \langle \chi \rangle f(c) n^{1/3},$$

$$\mathrm{Var}(L_{n,n}^{\mathrm{BM}}) \approx (\langle \chi^2 \rangle - \langle \chi \rangle^2) f^2(c) n^{2/3}, \tag{6}$$

where we have used the known exact values [5], $\langle \chi \rangle = -1.7711\cdots$ and $\langle \chi^2 \rangle - \langle \chi \rangle^2 = 0.8132\ldots$. The recursion relation (1) can also be viewed as a $(1+1)$-dimensional directed polymer problem [22,23] and some asymptotic results [such as the $O(n^{2/3})$ behavior of the variance of $L_{n,n}$ for large n] can be obtained using the arguments of universality [22]. However, the limiting Tracy-Widom distribution and the associated exact scale factors in Eqs. (4)–(6) derived in [4] cannot be obtained simply from the universality arguments.

As it has been mentioned above, the level structure depicted in Fig. 1(b) can be viewed as three-dimensional terraces. Namely, let us add one extra height dimension to the system of level lines in Fig. 1(b). Each time when we cross the level line constructed via the recursion algorithm (1), we increase the height by one. Hence, the length $L_{i,j}^{\mathrm{BM}}$ can be interpreted as the height of a surface above the two-dimensional (2D) (i,j) plane. Considering the 2D projection of the level lines separating the adjacent terraces in Fig. 1(b), we can note that the rule (1) prohibits the overlap of these level lines, i.e., different level lines cannot have common segments or edges—see Figs. 2(a) and 2(b). The resulting three-dimensional system of terraces shown in Fig. 2(b) is in one-to-one correspondence with the two-dimensional system of level lines in Fig. 2(a). Note that a similar (but not identical) model with terrace structures and the associated level lines also appeared in an anisotropic three-dimensional (3D) directed percolation model [24], which in turn is also related to the directed polymer problem studied by Johansson [25]. However, the levels lines in the directed percolation model can overlap. In contrast, the level lines in our model do not

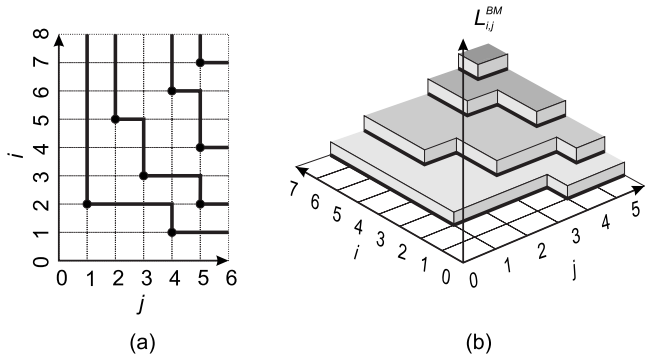

FIG. 2. Bernoulli matching model: (a) Level lines. This is just Fig. 1(b) rotated by 90°. Note that adjacent level lines have no common segments or edges; (b) 3D terracelike representation of the heights.

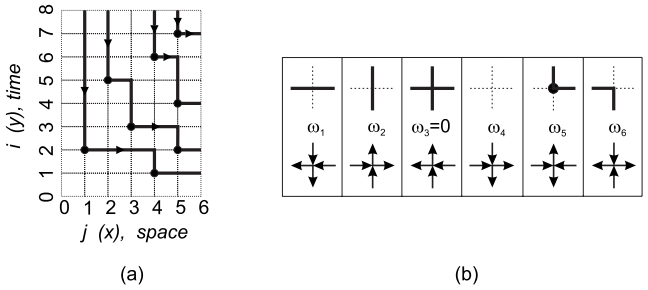

FIG. 3. (a) System of 2D level lines separating adjacent terraces and the left-hand corners depicted by big dots, (b) the weights of the associated vertices. Note that the labels on the axes of (a) correspond to the labels in Fig. 2(b), but the time actually runs vertically downwards. The labels (the numbers) on the vertical axis do not correspond to time.

overlap. These two models are nevertheless related by a nonlinear transformation [4,24]. The connections among all mentioned and some other related models is briefly discussed in the conclusion.

All previously derived results (4)–(6) deal with the height statistics in the three-dimensional terrace representation of the Bernoulli matching model. In terms of the level lines in the projected 2D plane [see Fig. 2(a)], the height $L_{i,j}^{\mathrm{BM}}$ is just the total number of level lines within the rectangle of side lengths i and j. Note that any configuration of the level lines in the 2D plane has an associated statistical weight (see later for details). These weights are such that one can interpret a projected 2D configuration of the BM model as a five-vertex model. This five-vertex model is an interesting model in its own right (apart from its connection to the BM model) and it is natural to study the statistical properties of various objects associated with this two-dimensional five-vertex model. One such quantity is the total number of level lines inside the rectangle with sides i and j that translates into the height in the BM model. Similarly, there are other random variables such as the total number of corners and the total number of horizontal segments in the rectangle of sides (i,j) that have not been studied before, and that have nontrivial and interesting statistical properties. For example, the left-hand corners shown by the big dotted points in Fig. 3(a) are the so-called nucleation centers for the terraces and play an important role in the mapping between the BM model and the so-called longest increasing subsequence (LIS) problem [4]. In the limit of large number of letters, c $(c\to\infty)$, one can show that these nucleation centers are Poisson distributed in the 2D plane with a uniform density $\rho_c=1/c$ [4]. However, for finite c, the statistics of the number of nucleation centers is, to our knowledge, still unknown. In particular, one would expect that even the average density $\rho_c(x,y)$ of the nucleation centers for finite c is nonuniform in the (x,y) plane and has a nontrivial form that reduces to the uniform value $\rho_c=1/c$ in the $c\to\infty$ limit. In the present paper, we shall calculate explicitly the average density $\rho_c(x,y)$ of the nucleation centers for finite c using the Bethe Ansatz technique.

The outline of this paper is as follows. In Sec. II, we explain the precise mapping between the Bernoulli matching model and the five-vertex model and we recall the associated Bethe equations. We then solve the Bethe equations in cylindrical geometry. This allows us in Sec. III to calculate the mean flux of the word lines in the five-vertex model as well as the associated large deviation function. Reverting to the Bernoulli matching model, our calculation leads to an independent derivation of the expectation of the longest common subsequence. In Sec. IV, we determine the statistics of the number of left-hand corners. In terms of the terrace model these left-hand corners play the role of nucleation centers: our approach allows us to calculate the mean number of such centers. Concluding remarks and a flowchart of various models related to the Bernoulli matching model are given in Sec. V.

II. BERNOULLI MATCHING AS A FIVE-VERTEX MODEL: BETHE EQUATIONS

The statistical weight of a projected 2D configuration of lines of the Bernoulli matching model depicted in Fig. 2 is the product of weights associated with the vertices on the 2D plane. Let $W(C)$ be the total weight of a full 2D configuration of the system of lines in Fig. 2(a). Note that a new terrace is nucleated with probability $p=1/c$ and hence we associate a weight p with each nucleation center. These nucleation centers are the left-hand corners of the lines shown by the big dotted points in Fig. 2(a). On the other hand, each empty vertex in Fig. 2(a) has a weight $q=1-p$ (corresponding to a mismatch in the BM model). The other filled vertices that are not left-hand corners have an associated weight of 1 each. Thus, we may write the total weight of a configuration C of vertices in Fig. 2(a) as

$$W(C)=p^{N_c}q^{N_e} \quad \text{with} \quad p=\frac{1}{c}, \tag{7}$$

where N_c and N_e are the numbers of left-hand corners [shown by big dotted points in Fig. 2(a)] and empty vertices, respectively.

One can view the configuration of trajectories in Fig. 2(a) as an assembly of world lines of hard-core particles moving on a one-dimensional lattice. In this alternative representation the horizontal (x) and the vertical (y) axes in Fig. 3(a)

are correspondingly the time and the space directions of the one-dimensional (1D) lattice gas picture. The dynamics of these hard-core particles is as follows. Particles are moving along the lines as indicated in Fig. 3 and each particle tries to jump to any of the empty slots available to it before the location of the next particle to its right. Let $P(s|m)$ denote the probability that a particle hops s steps, given that the next particle to its right is at a distance m. Thus, there are $(m-1)$ holes between the two particles and therefore $s=0,1,2,\ldots(m-1)$. Equation (7) for the total weight $W(C)$ of the configuration dictates the following choice of this hopping probability:

$$P(s|m) = p^{1-\delta_{s,0}} q^{m-1-s}. \tag{8}$$

One can easily check that the probability $P(s|m)$ is properly normalized, $P(k|m)=\Sigma_{k=0}^{m-1} P(k|m)=1$.

There are five types of possible vertices with nonzero weights as shown in Fig. 3(b). Since the level lines never cross each other, the weight ω_3 is always zero. As it is seen from definition of vertices in Fig. 3(b), we draw a solid line for the arrow $\rightarrow$ or $\downarrow$. Otherwise we leave a bond empty. These rules slightly differ from the standard notations of Baxter [26], but correspond to the notations used in Ref. [27] where the Bethe Ansatz equation has been considered for the five-vertex model. Suppose now that the weights $\{\omega_1,\ldots,\omega_6\}$ are as follows:

$$\omega_1 = 1, \quad \omega_2 = 1, \quad \omega_3 = 0, \quad \omega_4 = q = 1-p,$$

$$\omega_5 = p, \quad \omega_6 = 1. \tag{9}$$

Let us note that in all Bethe equations below the corner weights ω_5 and ω_6 always enter in the combination $\omega_5\omega_6$. Hence, only the product weight $\omega_5\omega_6=p$ is properly defined. Henceforth, we shall call the left-hand corners [shown by the big dotted points in Fig. 3(a)] as the corners on which we will mainly focus. We are not interested here in the right-hand corners that correspond to the termination of the horizontal part of each world line.

We first consider our vertex model in a cylindrical geometry with N sites in each row and we assume that there are infinite number of rows. As usual, the number m of up arrows $(\uparrow)$ in each row is conserved. Because of the one-to-one correspondence of up arrows and solid vertical bonds in a row, m is the number of solid vertical bonds and $N-m$ is the number of empty vertical bonds in a row.

The grand canonical partition function, $Z(p)$, of a five-vertex model reads as

$$Z(p) = \sum_{\text{conf}} \omega_1^{N_h} \omega_2^{N_v} \omega_4^{N_e} (\omega_5\omega_6)^{N_c}, \tag{10}$$

where N_h, N_v, N_e, N_c are correspondingly the numbers of horizontal and vertical bonds, empty vertices, and corners [by corners we mean only left-hand corners shown by the big dots in Fig. 3(a)] for any particular configuration of the lines (or equivalently of the arrows). The summation in (10) runs over all available configurations of these world lines, satisfying the noncrossing and nonoverlapping constraints.

The standard Bethe equation for the roots z_j of the five-vertex model is as follows (see [27] for details):

$$z_j^N = (-1)^{N-m-1} \prod_{i=1}^{N-m} \frac{1-\Delta z_j}{1-\Delta z_i} \quad (j=1,2,\ldots,N-m), \tag{11}$$

where

$$\Delta = \frac{\omega_1\omega_2 - \omega_5\omega_6}{\omega_2\omega_4}. \tag{12}$$

The highest eigenvalue, Λ_m, can be written in two equivalent forms,

$$\Lambda_m = \omega_1^{N-m} \prod_{j=1}^{m} \frac{\omega_5\omega_6}{\omega_1 z_j - \omega_4} + \omega_4^{N-m} \prod_{j=1}^{m} \frac{\omega_2\omega_1 z_j - \omega_2\omega_4 - \omega_5\omega_6 z_j}{\omega_1 z_j - \omega_4}$$

$$= \omega_2^m \omega_4^{N-m} \prod_{j=1}^{N-m} \left(1 + \frac{\omega_5\omega_6}{\omega_2\omega_4} z_j\right). \tag{13}$$

The second expression will be more convenient for our further computations. Given the weights (9), we find

$$\Delta = \frac{1-p}{q} = 1. \tag{14}$$

Hence, we obtain the following Bethe equations:

$$z_j^N = (-1)^{N-m-1} \prod_{i=1}^{N-m} \frac{1-z_j}{1-z_i}. \tag{15}$$

Λ_m is given by Eq. (13),

$$\Lambda_m = (1-q)^m \prod_{j=1}^{m} \frac{1}{z_j - q} + q^N \prod_{j=1}^{m} \frac{z_j - 1}{z_j - q} = \prod_{j=1}^{N-m} (1 + p z_j) \tag{16}$$

with z_j $(j=1,\ldots,N-m)$ defined by (15).

Equations (15) and (16) are almost identical to the Bethe equation for roots and to the expression for the highest eigenvalue of the transfer matrix of the totally asymmetric exclusion process (TASEP) [28]. In the context of the exclusion process, these equations have been studied by many authors [29–32] (for a recent review see [33]).

III. ANALYSIS OF THE BETHE EQUATIONS

A. Statistics of the world lines: The flux

The averaged flux, $\bar{\Phi}$, in the system of world lines shown in Fig. 3(a) is equal to the typical length (normalized per N) of a horizontal segment, $\langle N_h \rangle$, between left- and right-hand corners averaged over all configurations. Hence, to study the statistics of Φ, we add an extra weight e^{μ} to each horizontal bond and each essential corner. We thus define the partition function $Z(p,\mu)$ with the following collection of weights [compare to (9)]:

$$\omega_1 = e^{\mu}, \quad \omega_2 = 1, \quad \omega_3 = 0, \quad \omega_4 = q = 1 - p, \quad \omega_5\omega_6 = pe^{\mu}. \tag{17}$$

The mean value $\langle N_h \rangle$ can then be computed in a standard way

$$\bar{\Phi} \equiv \frac{\langle N_h \rangle}{N} = \frac{\sum_{\text{conf}} N_h (1-p)^{N_e} (e^{\mu})^{N_h} (pe^{\mu})^{N_c}}{N \sum_{\text{conf}} (1-p)^{N_e} (e^{\mu})^{N_h} (pe^{\mu})^{N_c}}$$

$$= \frac{1}{N} \frac{\partial}{\partial \mu} \ln Z(p,\mu)|_{\mu=0}. \tag{18}$$

The fluctuations of the average flux can be derived in the similar way,

$$\mathrm{Var}(\Phi) \equiv \frac{\langle N_h^2 \rangle}{N^2} - \frac{\langle N_h \rangle^2}{N^2} = \frac{1}{N^2} \frac{\partial^2}{\partial \mu^2} \ln Z(p,\mu)|_{\mu=0}. \tag{19}$$

The particular choice of weights (17) leads to the following expression for Δ defined in (12),

$$\Delta = \frac{\omega_1\omega_2 - \omega_5\omega_6}{\omega_2\omega_4} = e^{\mu}, \tag{20}$$

while the general form of the Bethe equation (11) remains unchanged. The highest eigenvalue Λ_n reads now as

$$\Lambda_m = \omega_2^m \omega_4^{N-m} \prod_{j=1}^{N-m} \left(1 + \frac{\omega_5\omega_6}{\omega_2\omega_4} z_j\right) = \prod_{j=1}^{N-m} (1 - p + pe^{\mu} z_j). \tag{21}$$

We proceed further using the technique of analyzing Bethe the equations proposed in [31]. Making the change of variables

$$y_j = z_j \Delta - 1 = z_j e^{\mu} - 1, \tag{22}$$

we can seek the solution of the Bethe equation (11) rewritten for y_j,

$$(y_j + 1)^{-N} y_j^Q = \Delta^{-N} (-1)^{Q-1} \prod_{i=1}^{Q} y_i, \tag{23}$$

in the form

$$y_j = B^{1/Q} e^{2\pi i j/Q} (1 + y_j)^{N/Q}, \tag{24}$$

where B is a constant that will be determined self-consistently, and $Q=N-m$. We arrive finally at the system of equations for the parametric determination of the partition function $Z(p,\mu)=(\Lambda_m)^N$ in the thermodynamic limit $N \to \infty$,

$$\frac{1}{N} \ln Z(p,\mu) = \sum_{j=1}^{Q} \ln(1 + py_j),$$

$$Q\mu = \sum_{j=1}^{Q} \ln(1 + y_j). \tag{25}$$

Equations (24) and (25) can be rewritten in a closed form using the standard residue formula—see, for example, Ref. [34],

$$\frac{1}{N} \ln Z(p,\mu) = \frac{1}{2\pi i} \oint \sum_{j=1}^{Q} \ln(1+py) \left(1 - \frac{N}{Q} \frac{y}{y+1}\right) \times \frac{dy}{y - B^{1/Q} e^{2\pi i j/Q} (1+y)^{N/Q}},$$

$$Q\mu = \frac{1}{2\pi i} \oint \sum_{j=1}^{Q} \ln(1+y) \left(1 - \frac{N}{Q} \frac{y}{y+1}\right) \times \frac{dy}{y - B^{1/Q} e^{2\pi i j/Q} (1+y)^{N/Q}}. \tag{26}$$

Solving (26) we express $\ln Z(p,\mu)$ as a series expansion

$$\frac{1}{N} \ln Z(p,\mu) = p \sum_{k=1}^{\infty} B^k \frac{\mathcal{F}(Nk,Qk)}{k},$$

$$\mu = \sum_{k=1}^{\infty} B^k \frac{(Nk-1)!}{(Qm)\,![(N-Q)k]!}, \tag{27}$$

where

$$\mathcal{F}(Nk,Qk) = \sum_{m'=0}^{Qk-1} \binom{Nk}{m'} (-p)^{Qk-m'-1}. \tag{28}$$

For the computation of the expectation and the variance of the flux it is sufficient to cut the series (28) at the second term,

$$\frac{1}{N} \ln Z(p,\mu) = pB\mathcal{F}(N,Q) + \frac{1}{2} pB^2 \mathcal{F}(2N,2Q), \tag{29a}$$

$$\mu = B \frac{(N-1)!}{Q\,!\,(N-Q)!} + B^2 \frac{(2N-1)!}{(2Q)\,!\,[2(N-Q)]!}, \tag{29b}$$

where

$$\mathcal{F}(N,Q) \simeq \binom{N}{Q-1} \frac{1}{1 + \frac{1-\rho}{\rho} p} = \frac{\rho}{p + \rho q} \binom{N}{Q-1},$$

$$\mathcal{F}(2N,2Q) \simeq \binom{2N}{2Q-1} \frac{1}{1 + \frac{1-\rho}{\rho} p} = \frac{\rho}{p + \rho q} \binom{2N}{2Q-1}, \tag{30}$$

and we have defined the density ρ (of world lines) in (30) in the limit $N \to \infty$ as

$$\rho = \frac{m}{N} \equiv \frac{N-Q}{N}. \tag{31}$$

Now we can extract the constant B from (29b) and substitute it into (29a). After some algebra and using the Stirling formula, we arrive at the expression for the free energy (mean value and fluctuations) of the system in the thermodynamic limit $N \to \infty$ in the ensemble with fixed density ρ of world (level) lines (i.e., with fixed total number of world lines, m

$=\rho N$) and fugacity, μ, of horizontal bonds (including corners),

$$\frac{1}{N}\ln Z_h(\rho,p,\mu) \equiv \frac{1}{N}\ln Z(p,\mu) = \frac{p(1-\rho)}{p+q\rho}\mu N + \frac{\sqrt{\pi}}{4}\frac{p}{p+q\rho}\frac{(1-\rho)^{3/2}}{\rho^{1/2}}\mu^2 N^{3/2}. \tag{32}$$

Differentiating Eq. (32) with respect to μ and setting $\mu=0$, we obtain the expectation of the flux $\bar{\Phi}$ in the system,

$$\bar{\Phi} = \frac{p(1-\rho)}{p+q\rho}. \tag{33}$$

This expression thus reproduces the result obtained in [24] by a different method. According to (19) the variance $\mathrm{Var}(\Phi)$ reads

$$\mathrm{Var}(\Phi) = \frac{\sqrt{\pi}}{4}\frac{p}{p+q\rho}\frac{(1-\rho)^{3/2}}{\rho^{1/2}}N^{-1/2}. \tag{34}$$

More generally, eliminating B between the two equations of (27), allows us to determine the moments of Φ to any desired order.

B. Expected length of the longest common subsequence in the Bernoulli matching model

Let us return to the system of world lines shown in Fig. 3(a). Our aim in this section is to derive a macroscopic coarse-grained equation of state for the average height $\bar{L}(x,y)$ in the BM model which is precisely the average length of the longest match. To derive this, we return to the system of world lines shown in Fig. 3(a), then we closely follow a similar line of arguments that were used in [24]. Let us first define $\partial_x L(x,y) \equiv \frac{\partial L(x,y)}{\partial x}$ and $\partial_y L(x,y) \equiv \frac{\partial L(x,y)}{\partial y}$. These are random variables and we would like to estimate their average values. First consider crossing the world lines by going horizontally along the x direction. The variable $\partial_x L(x,y)$ takes the values

$$\partial_x L(x,y) = \begin{cases} 0 & \text{if we do not cross a level line,} \\ 1 & \text{if we cross a level line.} \end{cases}$$

Thus the average value of $\partial_x L(x,y)$ is the number of lines encountered per unit distance along the horizontal (x) direction. In the 1D lattice gas language this is just the density ρ of particles. Hence, our first relation is

$$\partial_x \bar{L}(x,y) = \rho. \tag{35}$$

Let us now try to relate the second object $\partial_y \bar{L}(x,y)$ also to the density ρ. For this, let us imagine crossing the world lines along the vertical (y) axis. It is clear that one can write

$$\partial_y \bar{L}(x,y) = \partial_x \bar{L}(x,y)\frac{\partial x}{\partial y} = \rho\frac{\partial x}{\partial y} \tag{36}$$

and then replace $\frac{\partial x}{\partial y}$ by its coarse-grained value, namely,

$$\frac{\partial x}{\partial y} \approx \frac{\langle N_h\rangle}{N_l}, \tag{37}$$

where N_l is the total number of world lines or equivalently the total number of particles in the 1D lattice with N sites. The right-hand side of Eq. (37) is just the mean length of the total horizontal segment in each row per world line and hence is also the mean horizontal distance traveled by each particle in unit time (note that one unit of time corresponds to traversing one row in the vertical direction) and hence it represents the macroscopic value of $\frac{\partial x}{\partial y}$ on the left-hand side of Eq. (37). Now, writing $N_l=\rho N$, we immediately obtain from Eqs. (37) and (36),

$$\partial_y \bar{L}(x,y) = \frac{\langle N_h\rangle}{N} = \bar{\Phi} = \frac{p(1-\rho)}{p+q\rho}, \tag{38}$$

where we have used the expression of $\bar{\Phi}$ from Eq. (33).

Note that the density ρ in both Eqs. (35) and (38) is an unknown macroscopic variable, which however can be eliminated from the two equations. This then gives the desired equation for the average surface height,

$$p[1-\partial_x \bar{L}(x,y)-\partial_y \bar{L}(x,y)] = q\partial_x \bar{L}(x,y)\partial_y \bar{L}(x,y). \tag{39}$$

Solving (39), we find the average profile of the surface shown in Fig. 2(b),

$$\bar{L}(x,y) = \frac{2\sqrt{pxy}-p(x+y)}{q}. \tag{40}$$

Note that this result is valid only in the regime $px<y<x/p$. This is because at $y=x/p$, $\bar{L}(x,y)=x$, and hence $\rho=1$, already achieves its maximum value. Hence, in the regime $y>x/p$, $\bar{L}(x,y)$ sticks to its value $\bar{L}(x,y)=x$. Symmetry arguments show that for $y<px$, $\bar{L}(x,y)$ sticks to the value $\bar{L}(x,y)=y$. We note that this result was first derived in [21] using probabilistic methods. Setting $x=y=n$ in (40) and recalling that $p=1/c$, we arrive at the expression

$$\bar{L}(n,n) = \frac{2}{1+\sqrt{c}}n, \tag{41}$$

which coincides with the first term of Eq. (6) for the expectation $\langle L_{n,n}\rangle$.

IV. STATISTICS OF CORNERS

In this section, we compute the mean number of corners which play the role of nucleation centers in the terrace model. As mentioned above, the expression (40) is valid in the angle $px<y<x/p$, while outside this region the average surface height is linearly increasing along x for $y\geq x/p$ and along y for $y\leq px$. Hence, the complete expression for $\bar{L}(x,y)$ is as follows:

$$\bar{L}(x,y) = \begin{cases} \frac{1}{q}[2\sqrt{pxy} - p(x+y)] & \text{if } px < y < x/p, \\ x & \text{if } y \geq x/p, \\ y & \text{if } y \leq px. \end{cases} \tag{42}$$

The function $\bar{L}(x,y)$ is depicted in Fig. 4 for two different values of p.

Since the function $\bar{L}(x,y)$ is symmetric with respect to axis $y=x$, we can consider the sector $x \leq y$ only and extend the obtained results to the sector $x \geq y$ afterwards. The local density of lines, $\rho(x,y)$, is defined in the following way:

$$\rho(x,y) = \begin{cases} \frac{\partial \bar{L}(x,y)}{\partial x} = 1 & \text{if } x,\ y \text{ belong to the sector I, } y \geq \frac{x}{p}, \\ \frac{\partial \bar{L}(x,y)}{\partial x} = \frac{\sqrt{p}}{q}\left(\sqrt{\frac{y}{x}} - \sqrt{p}\right) & \text{if } x,\ y \text{ belong to the sector II, } x \leq y < \frac{x}{p}. \end{cases} \tag{43}$$

Note that in all the results that we derived in the five-vertex model in a cylindrical geometry we have assumed a constant density ρ of world lines using it as a given fixed parameter of the model. Now, as seen above in Eq. (43), for the BM model in 2D, the density of lines $\rho(x,y)$ is not a constant, but is a function of the space. To use the five-vertex results in calculating the mean corner density in the BM model, we will use a coarse-grained description in the following sense. We consider the BM model on a very big lattice. Now, we consider a part of these world lines over a sufficiently big coarse-grained region around the point (x,y). In this local region, we will consider the line density $\rho(x,y)$ to be sufficiently slowly varying function and use it as a constant input in the corresponding five-vertex model to calculate the local corner density in the region around the point (x,y).

For each value of the fixed line density, $\rho(x,y)$, we need to evaluate the corner density, $\rho_c(x,y)$ from the five-vertex model. The vertex weights in this case are as follows [compare to (9) and (17)]:

$$\omega_1 = 1, \quad \omega_2 = 1, \quad \omega_3 = 0, \quad \omega_4 = q = 1-p, \quad \omega_5\omega_6 = pe^{\alpha}. \tag{44}$$

The mean value $\langle N_c \rangle$ can be computed as follows:

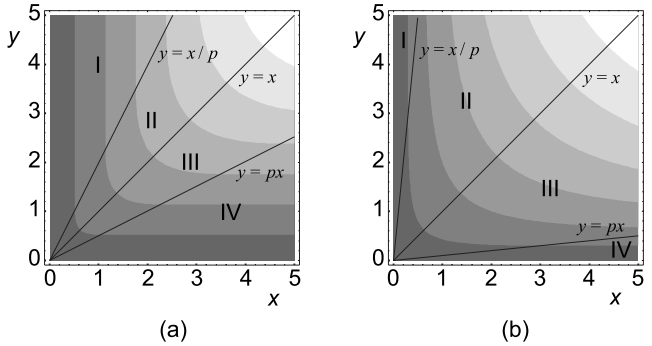

FIG. 4. Averaged surface height $\bar{L}(x,y)$ given by (42) for two different values: (a) $p=0.5$ and (b) $p=0.1$.

$$\langle N_c \rangle = \frac{\sum_{\text{conf}} N_c (1-p)^{N_e} (pe^{\alpha})^{N_c}}{\sum_{\text{conf}} (1-p)^{N_e} (pe^{\alpha})^{N_c}} = \frac{\partial}{\partial \alpha} \ln Z(p,\alpha)|_{\alpha=0}. \tag{45}$$

The fluctuations of the average number of corners can be derived in the similar way as the variance of the flux. Thus, the expression for the variance $\mathrm{Var}(N_c)$ is as follows:

$$\mathrm{Var}(N_c) \equiv \langle N_c^2 \rangle - \langle N_c \rangle^2 = \frac{\partial^2}{\partial \alpha^2} \ln Z(p,\alpha)|_{\alpha=0}. \tag{46}$$

The highest eigenvalue Λ_m (13) now reads as

$$\Lambda_n = \prod_{j=1}^{N-m} (1 - p + pe^{\mu} z_j) = \left(\frac{q}{1-pe^{\alpha}}\right)^{Q} \prod_{j=1}^{Q} (1 + pe^{\alpha} y_j) \tag{47}$$

with

$$y_j = z_j \Delta - 1 = z_j \frac{q}{1-pe^{\alpha}} - 1$$

and

$$\Delta = \frac{\omega_1\omega_2 - \omega_5\omega_6}{\omega_2\omega_4} = \frac{1-pe^{\alpha}}{q}. \tag{48}$$

Solving Bethe equations, we get the parametric system of equations defining the free energy [compare to (27)],

$$\frac{1}{N} \ln Z(p,\alpha) = -Q \ln \frac{1-pe^{\alpha}}{q} + pe^{\alpha} \sum_{k=1}^{\infty} B^k \frac{\mathcal{F}(Nk,Qk)}{k},$$

$$\ln \frac{1-pe^{\alpha}}{q} = \sum_{k=1}^{\infty} B^k \frac{(Nk-1)!}{(Qm)!\,[(N-Q)k]!}, \tag{49}$$

with $\mathcal{F}(Nk,Qk)$ as in (28). Proceeding as in Sec. III, we arrive at the following expression for the free energy of the system in the ensemble with fixed density, ρ, and fugacity of corners, α:

$$\frac{1}{N}\ln Z_c(\rho,p,\alpha) \equiv \frac{1}{N}\ln Z(p,\alpha) = -\frac{\rho(1-\rho)(1-pe^{\alpha})}{pe^{\alpha}+(1-pe^{\alpha})\rho}\eta N + \frac{\sqrt{\pi}}{2}\frac{pe^{\alpha}}{pe^{\alpha}+(1-pe^{\alpha})\rho}\frac{(1-\rho)^{3/2}}{\rho^{1/2}}\eta^2 N^{3/2}, \tag{50}$$

where

$$\eta = \ln\frac{1-pe^{\alpha}}{q}. \tag{51}$$

Using Eqs. (45), (50) and (51) we derive the expression for the corner density

$$\rho_c = \frac{\langle N_c\rangle}{N^2} = \frac{1}{N^2}\frac{\partial \ln Z_c(\rho,p,\alpha)}{\partial\alpha}\bigg|_{\alpha=0}. \tag{52}$$

After simple computations, we obtain for ρ_c the following equation:

$$\rho_c = \frac{\rho(1-\rho)p}{p+q\rho}, \tag{53}$$

where for ρ we should understand the local line density $\rho(x,y)$ given by (43). Substituting (43) into (53) we arrive at

$$\rho_c(x,y) = \begin{cases} 0 & \text{for } y \geq \dfrac{x}{p}, \\ \dfrac{p}{q^2}\dfrac{\left(\sqrt{\dfrac{y}{x}}-\sqrt{p}\right)\left(1-\sqrt{p}\sqrt{\dfrac{y}{x}}\right)}{\sqrt{\dfrac{y}{x}}} & \text{for } x \leq y < \dfrac{x}{p}. \end{cases} \tag{54}$$

Since the function $\rho_c(x,y)$ is symmetric with respect to $y=x$ axis, we can straightforwardly reconstruct the values of $\rho_c(x,y)$ in the regions III and IV from (54).

The average number, $\langle N_c\rangle$, of corners in the square box Ω of size $L\times L$ can be obtained by integrating the corner density, $\rho_c(x,y)$ in Ω where Ω denotes the regions II and III,

$$\langle N_c\rangle = \int_{\Omega}\rho_c(x,y)dxdy = 2\int_0^L dx\int_{px}^{x}\rho_c(x,y)dy = 2\int_0^L dx\int_{px}^{x}\frac{p}{q^2}\frac{\left(\sqrt{\dfrac{y}{x}}-\sqrt{p}\right)\left(1-\sqrt{p}\sqrt{\dfrac{y}{x}}\right)}{\sqrt{\dfrac{y}{x}}}dy = \frac{L^2 p(1-\sqrt{p})^3(3+\sqrt{p})}{3q^2}, \tag{55}$$

where $q=1-p$. It is easily seen that at $p\to 0$ the averaged number of corners, $\langle N_c\rangle$ tends to the value pL^2 that corresponds to the mean of the Poisson distributed corners or nucleation centers.

We conclude this section with the following remark. The definition of the density of level lines, ρ, deserves special attention. We distinguish the global, ρ_{gl}, and local, ρ_{loc}, densities, defined, respectively, as follows:

$$\rho_{\mathrm{gl}} = \frac{m}{N} \equiv \frac{\bar{L}(x,y)}{N}\bigg|_{x=y=N} = \frac{2\sqrt{p}}{1+\sqrt{p}}, \tag{56a}$$

$$\rho_{\mathrm{loc}} = \frac{\partial\bar{L}(x,y)}{\partial x}\bigg|_{x=y=N} = \frac{\sqrt{p}}{1+\sqrt{p}}. \tag{56b}$$

Note that Eqs. (33)–(35) are consistent with the density $\rho=\rho_{\mathrm{loc}}$ defined in (56b).

V. CONCLUSION

To summarize, in this paper we have shown how to use the Bethe Ansatz technique to compute the average number of lines and corners in a 2D five-vertex model that originated from the Bernoulli matching model of the alignment of two random sequences. The asymptotic result for the average number of lines in a square of size $(n\times n)$, which in the Bernoulli matching model is the average length $\langle L_{n,n}^{\mathrm{BM}}\rangle$ of the longest match between two random sequences each of length n, was already known by several alternate methods. These include a purely probabilistic method [21], the cavity method of the spin-glass physics [23], via mapping to a lattice gas of interacting particle systems [24] and also via a mapping [4] to the Johansson's directed polymer problem [25]. Here, we have provided yet another method namely the standard Bethe Ansatz technique to compute this asymptotic result. Moreover, we were also able to compute, by this method, the average density of corners in the five-vertex model. While the number of lines has an immediate physical meaning in the context of the sequence matching problem (namely, this is just the number of optimal matches between two random sequences), it is difficult to relate the corner density in the 2D plane to a directly measurable observable in the sequence matching problem. In an indirect way the corner density is related to the degeneracy of the optimal match. However, the corner density has a direct physical meaning in terms of the world lines of the totally asymmetric exclusion process. The corner density is just the total number of particle jumps per unit time and per unit length.

There are several models and the associated techniques that are directly or indirectly related to the Bernoulli matching model [6]. These models include the (1+1)-dimensional directed polymer problem in a random medium [25], an anisotropic (2+1)-dimensional directed percolation model [24], the longest increasing subsequence problem [35], a (1+1)-dimensional anisotropic ballistic deposition model [36], and also the two-dimensional five-vertex model described in this paper. They all share the common fact that there is a suitable random variable in the model whose limiting distribution is described by the Tracy-Widom law first appearing as the limit distribution of the largest eigenvalue of a random

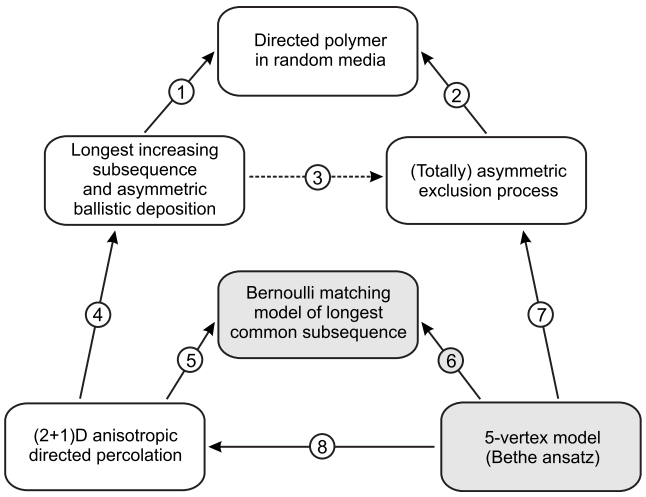

FIG. 5. A flowchart of the models related to the Bernoulli matching model of sequence alignment.

matrix drawn from the Gaussian unitary ensemble. In Fig. 5, we provide a flowchart of these different models and the connections between them are depicted by numbered arrows which designate the methods of solutions of listed problems. Below we briefly comment on these connections.

The first comprehensive solution for the distribution of the ground-state energy of directed polymer in random media (DPRM) has been obtained by Johansson in [25] by mapping this model to the longest increasing (nondecreasing) subsequence (LIS) in a random sequence of integers, known also as the Ulam problem (see arrow 1). He also discussed the possible connection between the asymmetric exclusion process (ASEP) and LIS but since this relation has not been deeply exploited, we have assigned to it a dashed arrow 3. The relation between ASEP and DPRM, arrow 2, is well known [37]. References [24] and [36] have offered the possibility for direct geometrical connection, shown by arrow 4 between the Tracy-Widom distribution of LIS and the scaled height in the anisotropic directed percolation (ADP) model. The relation between the Bernoulli matching model and the ADP model (arrow 5) is given by the nonlinear transform established in [4]. The solution of asymmetric exclusion process by Bethe Ansatz, shown by arrow 7 is a subject of many investigations (see [33] for references). As it has been mentioned in [24], the ADP model can be solved by mapping it to a five-vertex model (this link is shown by arrow 8) providing an alternative derivation (using Bethe Ansatz) of some results dealing with the statistics of LIS.

The shaded cells connected by arrow 6 constitute the subject of our current work—the use of the Bethe Ansatz technique for the two-dimensional five-vertex model. This thus adds one more standard technique of statistical physics to the number of existing methods that have already been used for investigation of problems depicted in Fig. 5. A challenging forthcoming goal would be the possibility to extract directly the Tracy-Widom distribution for the scaled height in the BM model using the Bethe Ansatz method. We note that the appearance of the Tracy-Widom law for the distribution of total current in a fragmentation model (which includes the TASEP) was explained before using the coordinate Bethe Ansatz method in [38]. This fragmentation model is similar to the particle hopping model considered in our work provided one exchanges the particles and holes. However, making a link between the total current distribution in the fragmentation model and the height distribution in the BM model is not so straightforward and remains an open question. Some recent attempts have been made in this direction in [39].

Finally let us note that the BM is a special case of a more general sequence alignment problem, where one includes a penalty (or cost) of mismatches [40]. Despite the fact that the recursion relation for the cost function can still be interpreted as a generalized directed polymer (DP) problem, the terrace-like three-dimensional structure (and hence the connection to the vertex models) in no longer valid [40,41]. It is not clear if one could use the Bethe Ansatz for this generalized DP problem.

The authors thank D. Dhar for many useful discussions. One of the authors (S.M.) acknowledges the support of the Indo-French Centre for the Promotion of Advanced Research (IFCPAR/CEFIPRA) under Contract No. 3404-2. One of the authors (S.N.) appreciates the partial support from Nouvelles Interfaces des Mathématiques (France) Grant No. ACI-NIM-2004-243.

[1] M. S. Waterman, *Introduction to Computational Biology* (Chapman and Hall, London, 1994).
[2] D. Gusfield, *Algorithms on Strings, Trees, and Sequences* (Cambridge University Press, Cambridge, 1997).
[3] R. Dubrin, S. Eddy, A. Krogh, and G. Mitchison, *Biological Sequence Analysis* (Cambridge University Press, Cambridge, 1998).
[4] S. N. Majumdar and S. Nechaev, Phys. Rev. E **72**, 020901(R) (2005).
[5] C. A. Tracy and H. Widom, Commun. Math. Phys. **159**, 151 (1994); see also *Proceedings of the ICM* (ICM, Beijing, 2002), Vol. I, p. 587.
[6] For a recent review of the appearance of Tracy-Widom distribution in several physics problems, see S. N. Majumdar, in Les Houches lecture notes, "Complex Systems," LXXXV (Elsevier, 2007), p. 179.
[7] S. B. Needleman and C. D. Wunsch, J. Mol. Biol. **48**, 443 (1970).
[8] T. F. Smith and M. S. Waterman, J. Mol. Biol. **147**, 195 (1981); Adv. Appl. Math. **2**, 482 (1981).
[9] M. S. Waterman, L. Gordon, and R. Arratia, Proc. Natl. Acad. Sci. U.S.A. **84**, 1239 (1987).
[10] S. F. Altschul *et al.*, J. Mol. Biol. **215**, 403 (1990).
[11] M. Zhang and T. Marr, J. Theor. Biol. **174**, 119 (1995).

[12] D. Sankoff and J. Kruskal, *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* (Addison Wesley, Reading, MA, 1983).
[13] A. Apostolico and C. Guerra, Alogorithmica **2**, 315 (1987).
[14] R. Wagner and M. Fisher, J. Assoc. Comput. Mach. **21**, 168 (1974).
[15] V. Chvátal and D. Sankoff, J. Appl. Probab. **12**, 306 (1975).
[16] J. Deken, Discrete Math. **26**, 17 (1979).
[17] J. M. Steele, SIAM J. Appl. Math. **42**, 731 (1982).
[18] V. Dancik and M. Paterson, *STACS94, Lecture Notes in Computer Science* (Springer, New York, 1994), Vol. 775, p. 306.
[19] K. S. Alexander, Ann. Appl. Probab. **4**, 1074 (1994).
[20] M. Kiwi, M. Loebl, and J. Matoušek, Advances in Mathematics **197**, 480 (2005).
[21] T. Seppäläinen, Ann. Appl. Probab. **7**, 886 (1997).
[22] T. Hwa and M. Lassig, Phys. Rev. Lett. **76**, 2591 (1996); R. Bundschuh and T. Hwa, Discrete Appl. Math. **104**, 113 (2000).
[23] J. Boutet de Monvel, Eur. Phys. J. B **7**, 293 (1999); Phys. Rev. E **62**, 204 (2000).
[24] R. Rajesh and D. Dhar, Phys. Rev. Lett. **81**, 1646 (1998).
[25] K. Johansson, Commun. Math. Phys. **209**, 437 (2000).
[26] R. J. Baxter, *Exactly Solvable Models in Statistical Mechanics* (Academic, London, 1989).
[27] J. D. Noh and D. Kim, Phys. Rev. E **49**, 1943 (1994).
[28] O. Golinelli and K. Mallick, J. Phys. A **40**, 5795 (2007).
[29] D. Dhar, Phase Transitions **9**, 51 (1987).
[30] L.-H. Gwa and H. Spohn, Phys. Rev. A **46**, 844 (1992).
[31] B. Derrida and J. L. Lebowitz, Phys. Rev. Lett. **80**, 209 (1998); B. Derrida and C. Appert, J. Stat. Phys. **94**, 1 (1999).
[32] O. Golinelli and K. Mallick, J. Phys. A **37**, 3321 (2004); **39**, 10647 (2006).
[33] O. Golinelli and K. Mallick, J. Phys. A **39**, 12679 (2006).
[34] M. Ablowitz, *Complex Variables: Introduction and Applications* (Cambridge University Press, Cambridge, 1997).
[35] S. M. Ulam, in *Modern Mathematics for the Engineers*, edited by E. F. Beckenbach (McGraw-Hill, New York, 1961), p. 261; J. M. Hammersley, *Proceedings of the VIth Berkeley Symposium on Mathematical Statistics and Probability* (University of California, Berkeley, 1972), Vol. 1, p. 345; A. M. Vershik and S. V. Kerov, Sov. Math. Dokl. **18**, 527 (1977); for a review, see D. Aldous and P. Diaconis, Bull. Am. Math. Soc. **36**, 413 (1999).
[36] S. N. Majumdar and S. Nechaev, Phys. Rev. E **69**, 011103 (2004).
[37] T. Halpin-Healy and Y.-C. Zhang, Phys. Rep. **254**, 215 (1995).
[38] A. Rakos and G. M. Schütz, J. Stat. Phys. **118**, 511 (2005).
[39] V. B. Priezzhev (private communication).
[40] R. Bundschuh and T. Hwa, Discrete Appl. Math. **104**, 113 (2000).
[41] M. V. Tamm and S. K. Nechaev, e-print arXiv:0710.4332.