

## Spanning trees in two dimensions

S. S. Manna\*

*Department of Mathematics, Yale University, New Haven, Connecticut 06520*Deepak Dhar<sup>†</sup> and S. N. Majumdar<sup>‡</sup>*Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400005, India*

(Received 17 June 1992)

We describe the results of an analytical and numerical study of the geometrical properties of random spanning trees on a square lattice. We determine exactly the probability distribution of the coordination number at a site on a random spanning tree. We argue that the probability that  $s$  sites get disconnected from the tree on deleting a bond at random from the tree varies as  $s^{-11/8}$  for large  $s$ . The probability that a loop of perimeter  $l$  is formed on adding an additional link at random varies as  $l^{-8/5}$  for large  $l$ . These distributions are also determined numerically in a Monte Carlo simulation on random spanning trees generated by using Broder's algorithm. The numerical results are in complete agreement with the theoretical predictions.

PACS number(s): 05.40.+j, 05.20.-y

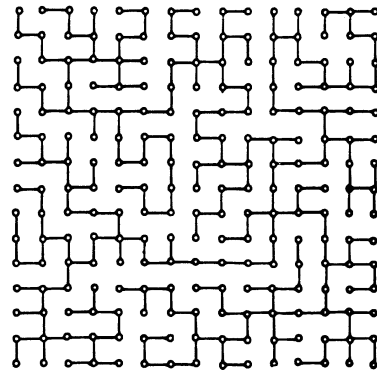
The problem of spanning trees is well known in statistical physics. It was shown by Kirchhoff that the spanning trees problem is related to the problem of determining the effective resistance between two nodes of a resistor network [1]. Fortuin and Kasteleyn showed that it is related to the  $q \rightarrow 0$  limit of the  $q$ -state Potts model [2]. More recent work on this problem has been reviewed by Wu [3]. Most of these results are related to the *number* of spanning trees on different graphs. However, the geometrical structure of the spanning trees has not been studied much so far. Recently we have established equivalence of the Abelian sandpile model [4] and the loop-erased self-avoiding walk problem [5] to the problem of spanning trees.

In this paper, we determine the exact probabilities of different coordination numbers at a given site on a random spanning tree on a square lattice. Let  $l$  denote the perimeter, and  $s$  denote the area of the loop formed on adding an extra link at a random position on the spanning tree. The corresponding probability distributions are denoted by  $\text{Prob}(l=l_0)$  and  $\text{Prob}(s=s_0)$ , respectively. The spanning tree problem on a square lattice is self-dual. Adding a link at random to a tree corresponds to deleting a link at random from the dual tree. Thus the probability  $\text{Prob}(s=s_0)$  also equals the probability that on deleting a link at random from a spanning tree, exactly  $s_0$  sites get disconnected. We determine exactly  $\text{Prob}(s=s_0)$  for  $s_0=1,2,3$  and  $\text{Prob}(l=l_0)$  for  $l_0=4$  and 6. We argue that  $\text{Prob}(s=s_0)$  varies as  $s_0^{-11/8}$  for large  $s_0$ , and  $\text{Prob}(l=l_0)$  varies as  $l_0^{-8/5}$  for large  $l_0$ . These conclusions are verified in a Monte Carlo simulation of the random spanning trees using Broder's algorithm [6].

A spanning tree on a graph of  $N$  sites is a set of  $N-1$  bonds which form a single connected cluster (see Fig. 1). The total number of distinct spanning trees on a finite graph is given by the well-known matrix tree theorem [7]. In the random spanning tree problem, all possible spanning trees are assumed to be equally likely. The coordina-

tion number of a site on a tree is the number of occupied edges meeting at that site. For a square lattice, it takes integer values from 1 to 4. The chemical path between two sites of a spanning tree is the unique path connecting them along the edges of the tree and is known to have fractal dimension  $5/4$  for random spanning trees in two dimensions [8,9]. For a given configuration of a spanning tree, we use a parallel-update burning process in which all sites having only one edge to an unburnt neighbor at the  $t$ th update are burnt (deleted) at the  $(t+1)$ th update. The burning time of a site is defined as the number of updates needed for the site to be burnt in this burning process. The sites corresponding to burning time 1 are called the leaves, and the site(s) corresponding to maximum burning time is (are) called the root(s) of the tree. A site  $i$  is said to be the descendent of site  $j$ , and  $j$  is called a predecessor of  $i$ , if and only if the unique path to  $i$  along the tree from a root site goes through  $j$ .

For definiteness, let us consider an  $L \times L$  square lattice with periodic boundary conditions. The sites are labeled  $1, 2, 3, \dots, L^2$ . For this lattice, we construct an  $L^2 \times L^2$  matrix  $\Delta$  as follows:

FIG. 1. A spanning tree on a  $15 \times 15$  square lattice.

$$\begin{aligned} \Delta_{ii} &= 4, \text{ for all } i = 1, \dots, L^2, \\ \Delta_{ij} &= -1, \text{ if } i \text{ and } j \text{ are nearest neighbors,} \\ \Delta_{ij} &= 0, \text{ otherwise.} \end{aligned}$$

Then by the matrix tree theorem [7], the number of spanning trees  $N_s$  on the lattice is given by any of the cofactors of matrix  $\Delta$  (all are equal)

$$N_s = \text{Cof}(\Delta).$$

It is straightforward to calculate the marginal probabilities of occurrence of given subgraphs in a randomly selected spanning tree by using the break-collapse method [2] and the matrix tree theorem. For example, to find the probability that in a random tree, all the edges of a given set  $A$  occur, and none of the edges of a given set  $B$  occur [to be denoted by  $\text{Prob}(A\bar{B})$  in the following], we simply construct a new graph  $L'$  from the original graph  $L$  by deleting the edges  $B$  and collapsing the edges  $A$ . Then if  $\Delta'$  is the matrix associated to  $L'$ , we get

$$\text{Prob}(A\bar{B}) = \frac{\text{Cof}(\Delta')}{N_s}.$$

For example, in Fig. 2, the probability  $f_1$ , that a randomly chosen site  $O'$  has coordination number 1, is the sum of four terms of the type  $\text{Prob}(e_4\bar{e}_5\bar{e}_6\bar{e}_7)$ . By symmetry these are equal. Hence

$$f_1 = 4\text{Prob}(e_4\bar{e}_5\bar{e}_6\bar{e}_7).$$

But  $\text{Prob}(e_4\bar{e}_5\bar{e}_6\bar{e}_7)$  has been shown to be equal to the concentration of sites with sandpile height 1 in the Abelian sandpile model on the square lattice, and is known [10]. Using this result, we get

$$f_1 = \frac{8}{\pi^2} \left( 1 - \frac{2}{\pi} \right) \approx 0.29454.$$

The probabilities  $f_2, f_3,$  and  $f_4$  that  $O$  has coordination

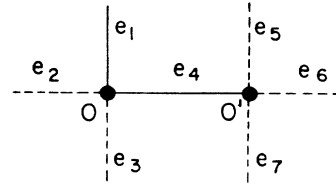


FIG. 2. A part of a large spanning tree. On deleting the edge  $e_1$ , the two sites  $O$  and  $O'$  get disconnected from the rest of the tree.

numbers 2, 3, and 4, respectively, are similarly calculated

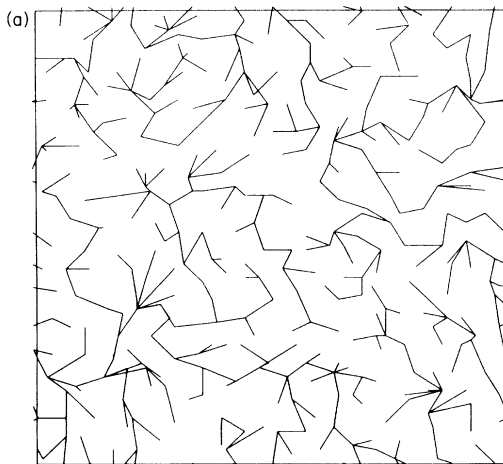
$$\begin{aligned} f_2 &= 4\text{Prob}(e_1e_2\bar{e}_3\bar{e}_4) + 2\text{Prob}(e_1e_3\bar{e}_2\bar{e}_4) \\ &= \frac{8}{\pi} - \frac{36}{\pi^2} + \frac{48}{\pi^3} \approx 0.44699, \end{aligned}$$

$$f_3 = 4\text{Prob}(e_1e_2e_3\bar{e}_4) = 2 - \frac{16}{\pi} + \frac{48}{\pi^2} - \frac{48}{\pi^3} \approx 0.22239,$$

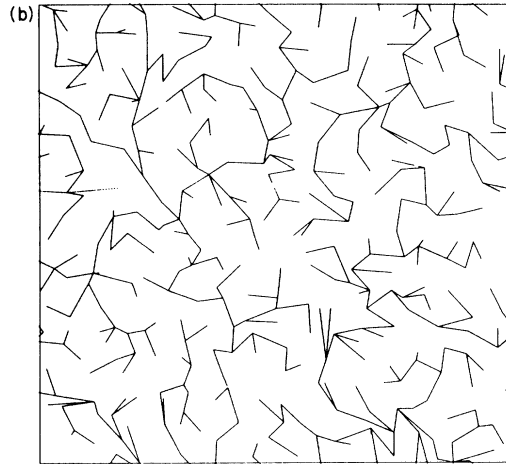
$$f_4 = 1 - f_1 - f_2 - f_3 = -1 + \frac{8}{\pi} - \frac{20}{\pi^2} + \frac{16}{\pi^3} \approx 0.03608.$$

Note that unlike  $f_1$ , the probabilities  $f_2, f_3,$  and  $f_4$  are not related directly to the concentrations of sites with heights 2, 3, or 4 in the sandpile model on the square lattice. The probabilities of some larger subgraphs can be calculated similarly. For example, the probability that on deleting a randomly selected edge  $e_1$  of the tree, only two sites get disconnected from the tree, is seen to be a sum of six equal terms of the form  $\text{Prob}(e_1e_4\bar{e}_2\bar{e}_3\bar{e}_5\bar{e}_6\bar{e}_7)/\text{Prob}(e_1)$  (Fig. 2). Evaluating the corresponding determinant we get

$$\begin{aligned} \text{Prob}(s=2) &= 12 \left[ \frac{9}{32} - \frac{9}{2\pi} + \frac{47}{2\pi^2} - \frac{48}{\pi^3} + \frac{32}{\pi^4} \right] \\ &\approx 0.12409. \end{aligned}$$



L = 100, T = 10



L = 303, T = 40

FIG. 3. Self-similarity of spanning trees under decimation. Only sites whose burning times are exact multiples of an integer  $\tau$  are kept. (a) A decimated tree on a  $100 \times 100$  lattice for  $\tau = 10$ . (b) Decimated tree on a  $303 \times 303$  lattice for  $\tau = 40$ . The magnification of (b) has been reduced by a factor  $4^{4/5}$  compared to (a).

In a similar way,  $\text{Prob}(s = 3)$  can be exactly computed as a rational polynomial of degree 7 in  $\pi^{-1}$  and we find (the determinants involved here also have been computed in Ref. [10], again in the context of the Abelian sandpile model, where details of calculation may be found)

$$\text{Prob}(s = 3) \approx 0.06575.$$

From the geometry of square lattice, it is clear that  $\text{Prob}(l = 4)$  is equal to  $\text{Prob}(s = 1) = f_1$ , and  $\text{Prob}(l = 6)$  is the same as  $\text{Prob}(s = 2)$ .

The calculation of these probabilities for larger  $s$  or  $l$  is straightforward, but tedious. These distributions are expected to have power-law tails:  $\text{Prob}(l) \sim l^{-x}$  and  $\text{Prob}(s) \sim s^{-y}$  for large  $l$  and  $s$ . The values of the exponents  $x$  and  $y$  can be deduced from the following renor-

malization argument: For any spanning tree configuration  $T$ , we may construct a decimated tree  $T'$ , where the burning times are rescaled by an integer factor  $\tau$  by deleting all sites whose burning times are not exact multiples of  $\tau$ . The remaining sites are connected to each other to form a spanning tree  $T'$  on the undeleted sites such that for all sites  $i$  and  $j$ ,  $i$  is a predecessor of  $j$  on  $T'$  if and only if  $i$  is a predecessor of  $j$  on  $T$ . As the distribution of burning times has power-law tails, the decimated trees corresponding to different values of  $\tau$  are expected to be statistically similar to each other for large  $\tau$  except for an overall scale factor. In Fig. 3, we show two decimated trees corresponding to  $\tau = 10$  on a  $100 \times 100$  lattice, and  $\tau = 40$  on a  $303 \times 303$  lattice. The pictures are consistent with the assumed statistical self-similarity of spanning trees under decimation.

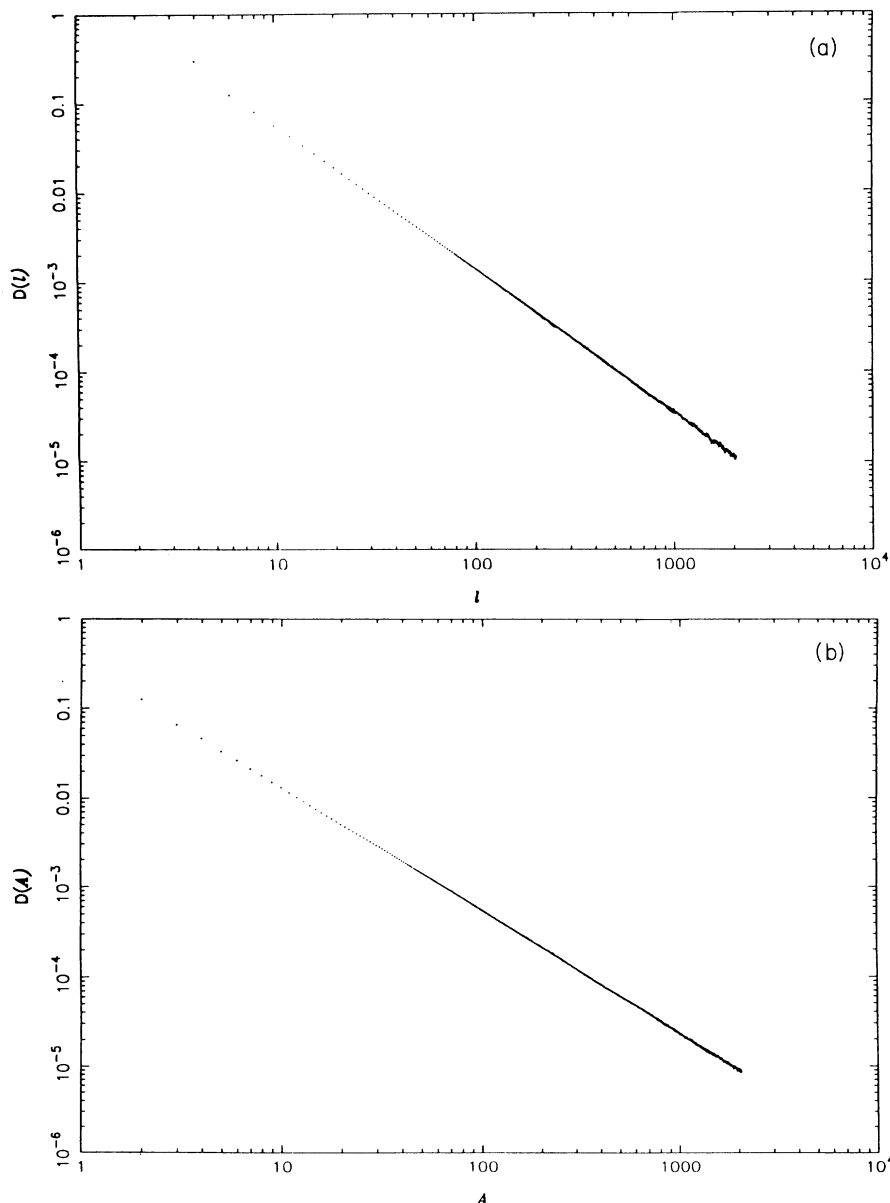


FIG. 4. The distribution of loop sizes on adding a bond at random to a spanning tree. (a) Double-logarithmic plot of the probability  $D(l)$  of forming a loop of perimeter  $l$  vs  $l$ . (b) Double-logarithmic plot of the probability  $D(A)$  of forming a loop of area  $A$  vs  $A$ .

Let  $b$  be the linear rescaling factor corresponding to rescaling of burning time by  $\tau$ . There are  $\tau$  links between two adjacent undecimated notes. As the fractal dimension of chemical paths on spanning trees is  $5/4$ , the average Euclidean distance between undecimated nodes scales as  $\tau^{4/5}$ . But this must equal the rescaling factor, giving us  $b \sim \tau^{4/5}$ . The set of sites whose burning time  $\geq \tau$  consists of the undecimated sites and those which lie on links between undecimated sites. Hence the fractional number of such sites is  $b^{-2}\tau \sim \tau^{-3/5}$ . Once this burning time distribution is known, it is easy to compute the perimeter distribution of the loop formed on adding an extra link on the spanning tree. We notice that the perimeter of the loop is also the perimeter of the connected cluster of sites that get disconnected from the dual tree on deleting the dual link. This cluster is compact and its perimeter typically scales as the burning time, *on the dual tree*, of the site in the disconnected cluster at the end of the deleted edge. Hence we conclude that

$$\text{Prob}(l \geq l_0) \sim l_0^{-3/5}. \quad (1)$$

A typical loop of perimeter  $l$  has linear size  $r \sim l^{4/5}$ . Since it is compact, its area  $\sim r^2$ . This implies that  $s \sim l^{8/5}$ . From Eq. (1), this implies that

$$\text{Prob}(s \geq s_0) \sim s_0^{-3/8} \quad (2)$$

so that we get  $x = \frac{8}{5}$  and  $y = \frac{11}{8}$ .

These results have been verified numerically by studying the statistics of spanning trees generated in a Monte Carlo simulation using Broder's algorithm [6]. We used an  $L \times L$  square lattice closed in all directions using helical boundary conditions. A random walker starts at the origin and performs an unbiased nearest-neighbor random walk. The walk is continued until all the sites of the lattice are visited at least once. We construct a tree by keeping only the bonds that correspond to the last exit of the walker from different lattice sites. There are  $L^2 - 1$  such

bonds as no exit bond from the end point of the walk is included. These occupied bonds form a spanning tree called the backward spanning tree. We treat spanning trees at time separated by  $3L^2$  time steps as essentially independent as this time interval for  $L \sim 10^3$  corresponds to updating more than  $\frac{2}{3}$  of the sites on the average. In our simulations, we generated 1530, 1479, 382, and 85 distinct spanning trees for  $L = 256, 512, 1024,$  and  $2048$ , respectively. The whole calculation took about 10 days of CPU on an IBM 6000 computer.

For each tree so generated, we determined the fractional number of sites having coordination numbers from 1 to 4. These fractions vary very little between different realizations of trees, and do not depend on  $L$  to the accuracy of our simulations. We get  $f_1 = 0.29454$ ,  $f_2 = 0.44700$ ,  $f_3 = 0.22239$ , and  $f_4 = 0.03607$ . The numerically determined values of  $\text{Prob}(s = s_0)$  for  $s_0 = 1, 2, 3$ , are  $0.29454$ ,  $0.12411$ , and  $0.06576$ . These are averages over approximately  $3.8 \times 10^8$  sites and the statistical error is less than 1 part in  $10^4$ . Clearly the agreement with theoretical predictions is excellent. For each tree generated, we calculated the distribution function  $\text{Prob}(s)$  by determining the number of sites that get disconnected on deleting an occupied bond, for all  $L^2 - 1$  possible choices of the bond to be deleted. Similarly, we calculated  $\text{Prob}(l)$ , by determining the perimeter of the loop formed on adding an extra bond for all the  $L^2 + 1$  possible choices of the bond to be added. These results are shown in Figs. 4(a) and 4(b). From the numerically determined slopes of these curves, we estimate that  $x = 1.600 \pm 0.002$  and  $y = 1.374 \pm 0.002$ . These values are in perfect agreement with our theoretical predictions.

*Note added.* After completion of this work, we received a copy of unpublished work from Burton and Pemantle in which they also have calculated the distribution of the coordination numbers of sites in spanning trees on the square lattice [11].

\*Present address: Department of Physics, Indian Institute of Technology, Powai, Bombay 400076, India.

†Electronic address: ddhar@tifrvax.bitnet.

‡Present address: AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, NJ 07974-0636.

[1] G. Kirchhoff, *Ann. Phys. Chem.* **72**, 497 (1847).

[2] C. M. Fortuin and P. W. Kasteleyn, *Physica* **57**, 536 (1972).

[3] F. Y. Wu, *Rev. Mod. Phys.* **54**, 235 (1982).

[4] S. N. Majumdar and D. Dhar, *Physica A* **185**, 129 (1992).

[5] S. N. Majumdar, *Phys. Rev. Lett.* **68**, 2329 (1992).

[6] A. Z. Broder, in *Proceedings of the 30th Annual IEEE Symposium on Foundations of Computer Science*, (IEEE, New York, 1989), p. 442.

[7] F. Harary, *Graph Theory* (Addison-Wesley, Reading, MA, 1990), p. 192.

[8] B. Duplantier, *J. Stat. Phys.* **54**, 581 (1989).

[9] A. Coniglio, *Phys. Rev. Lett.* **62**, 3054 (1989).

[10] S. N. Majumdar and D. Dhar, *J. Phys. A* **24**, L357 (1991).

[11] R. Burton and R. Pemantle (unpublished).