

---

# What is Percolation?

---

## 1.1 Modelling a Random Medium

Suppose we immerse a large porous stone in a bucket of water. What is the probability that the centre of the stone is wetted? In formulating a simple stochastic model for such a situation, Broadbent and Hammersley (1957) gave birth to the ‘percolation model’. In two dimensions their model amounts to the following. Let  $\mathbb{Z}^2$  be the plane square lattice and let  $p$  be a number satisfying  $0 \leq p \leq 1$ . We examine each edge of  $\mathbb{Z}^2$  in turn, and declare this edge to be *open* with probability  $p$  and *closed* otherwise, independently of all other edges. The edges of  $\mathbb{Z}^2$  represent the inner passageways of the stone, and the parameter  $p$  is the proportion of passages which are broad enough to allow water to pass along them. We think of the stone as being modelled by a large, finite subsection of  $\mathbb{Z}^2$  (see Figure 1.1), perhaps those vertices and edges of  $\mathbb{Z}^2$  contained in some specified connected subgraph of  $\mathbb{Z}^2$ . On immersion of the stone in water, a vertex  $x$  inside the stone is wetted if and only if there exists a path in  $\mathbb{Z}^2$  from  $x$  to some vertex on the boundary of the stone, using open edges only. Percolation theory is concerned primarily with the existence of such ‘open paths’.

If we delete the closed edges, we are left with a random subgraph of  $\mathbb{Z}^2$ ; we shall study the structure of this subgraph, particularly with regard to the way in which this structure depends on the numerical value of  $p$ . It is not unreasonable to postulate that the fine structure of the interior passageways of the stone is on a scale which is negligible when compared with the overall size of the stone. In such circumstances, the probability that a vertex near the centre of the stone is wetted by water permeating into the stone from its surface will behave rather similarly to the probability that this vertex is the endvertex of an infinite path of open edges in  $\mathbb{Z}^2$ . That is to say, the large-scale penetration of the stone by water is related to the existence of infinite connected clusters of open edges.

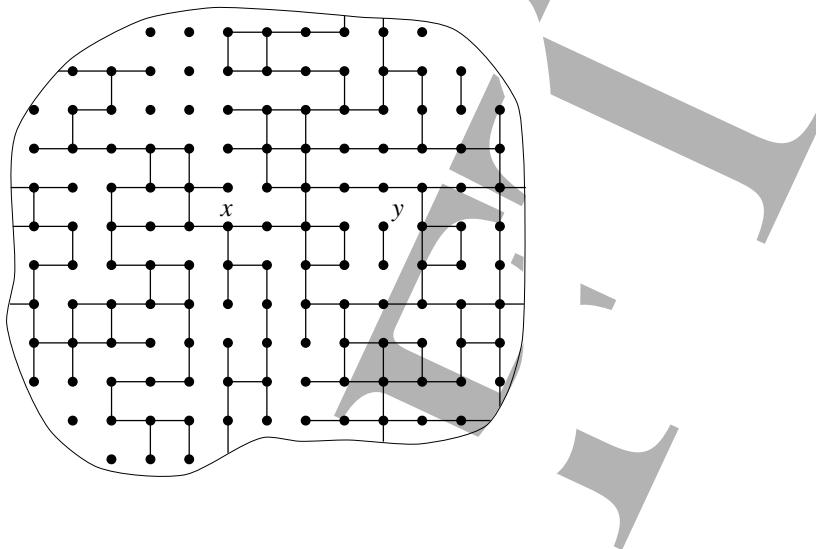


Figure 1.1. A sketch of the structure of a two-dimensional porous stone. The lines indicate the open edges; closed edges have been omitted. On immersion of the stone in water, vertex  $x$  will be wetted by the invasion of water, but vertex  $y$  will remain dry.

When can such infinite clusters exist? Simulations are handy indicators of the likely structure of the lattice, and Figure 1.2 contains such pictures for four different values of  $p$ . When  $p = 0.25$ , the connected clusters of open edges are isolated and rather small. As  $p$  increases, the sizes of clusters increase also, and there is a critical value of  $p$  at which there forms a cluster which pervades the entire picture. In loose terms, as we throw in more and more open edges, there comes a moment when large-scale connections are formed across the lattice. The pictures in Figure 1.2 are of course finite. If we were able to observe the whole of the infinite lattice  $\mathbb{Z}^2$ , then we would see that all open clusters are finite when  $p$  is small, but that there exists an infinite open cluster for large values of  $p$ . In other words, there exists a critical value  $p_c$  for the edge-density  $p$  such that all open clusters are finite when  $p < p_c$ , but there exists an infinite open cluster when  $p > p_c$  (such remarks should be interpreted ‘with probability 1’). Drinkers of Pernod are familiar with this type of phenomenon—the transparency of a glass of Pernod is undisturbed by the addition of a small amount of water, but in the process of adding the water drop by drop, there arrives an instant at which the mixture becomes opaque.

The occurrence of a ‘critical phenomenon’ is central to the appeal of percolation. In physical terms, we might say that the wetting of the stone is a ‘surface effect’ when the proportion  $p$  of open edges is small, and a ‘volume effect’ when  $p$  is large.

The above process is called ‘bond percolation on the square lattice’, and it is

the most studied to date of all percolation processes. It is a very special process, largely because the square lattice has a certain property of self-duality which turns out to be extremely valuable. More generally, we begin with some periodic lattice in, say,  $d$  dimensions together with a number  $p$  satisfying  $0 \leq p \leq 1$ , and we declare each edge of the lattice to be open with probability  $p$  and closed otherwise. The resulting process is called a ‘bond’ model since the random blockages in the lattice are associated with the *edges*. Another type of percolation process is the ‘site’ percolation model, in which the *vertices* rather than the edges are declared to be open or closed at random, the closed vertices being thought of as junctions which are blocked to the passage of fluid. It is well known that every bond model may be reformulated as a site model on a different lattice, but that the converse is false (see Section 1.6). Thus site models are more general than bond models. They are illustrated in Figure 1.9.

We may continue to generalize in several directions such as (i) ‘mixed’ models, in which both edges and vertices may be blocked, (ii) inhomogeneous models, in which different edges may have different probabilities of being open, (iii) long-range models, in which direct flow is possible between pairs of vertices which are very distant (in the above formulation, this may require a graph with large or even infinite vertex degrees), (iv) dependent percolation, in which the states of different edges are not independent, and so on. Mathematicians have a considerable talent in the art of generalization, and this has not been wasted on percolation theory. Such generalizations are often of considerable mathematical and physical interest; we shall however take the opposite route in this book. With few exceptions, we shall restrict ourselves to bond percolation on the  $d$ -dimensional cubic lattice  $\mathbb{Z}^d$  where  $d \geq 2$ , and the main reasons for this are as follows. As the level of generality rises, the accessibility of results in percolation theory is often diminished. Arguments which are relatively simple to explain in a special case can become concealed in morasses of geometrical and analytical detail when applied to some general model. This is not always the case, as illustrated by the proofs of exponential decay when  $p < p_c$  (see Chapter 5) and of the uniqueness of the infinite open cluster when it exists (see Chapter 8). It is of course important to understand the limitations of an argument, but there may also be virtue in trying to describe something of the theory when stripped of peripheral detail. Bond percolation on  $\mathbb{Z}^d$  is indeed a special case, but probably it exhibits the majority of properties expected of more general finite-range percolation-type models.

## 1.2 Why Percolation?

As a model for a disordered medium, percolation is one of the simplest, incorporating as it does a minimum of statistical dependence. Its attractions are manyfold. First, it is easy to formulate but not unrealistic in its qualitative predictions for random media. Secondly, for those with a greater interest in more complicated processes, it is a playground for developing mathematical techniques and insight.

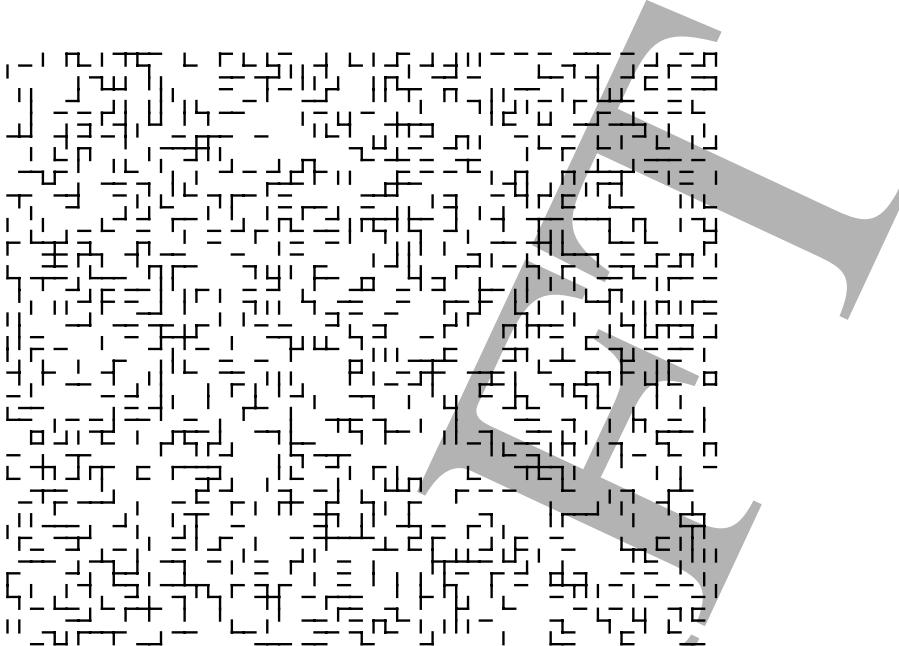
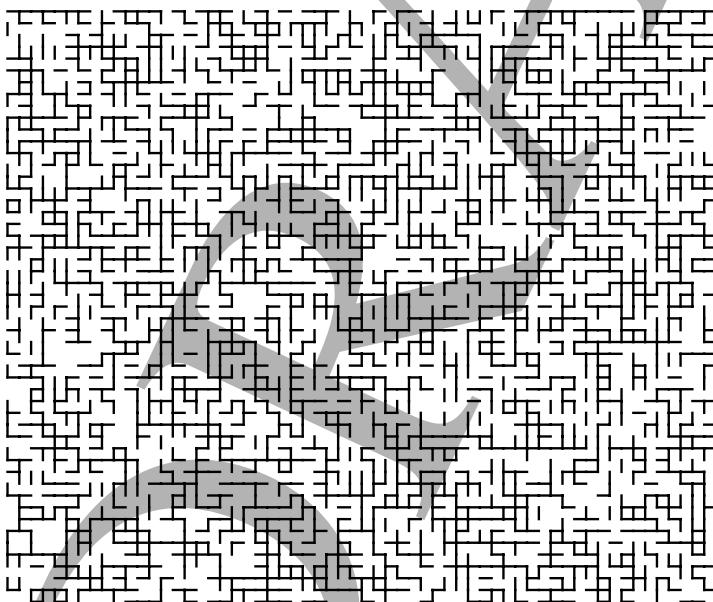
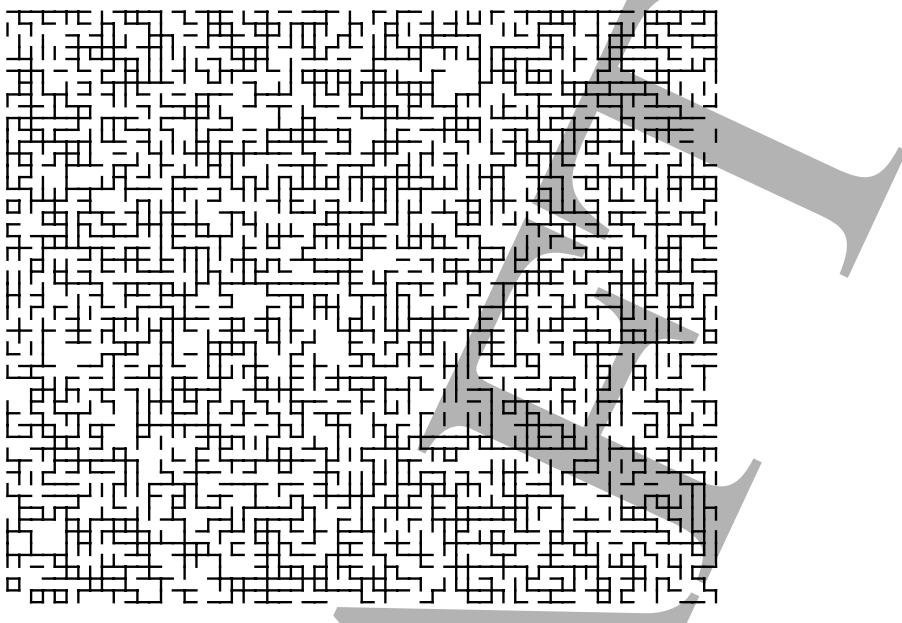
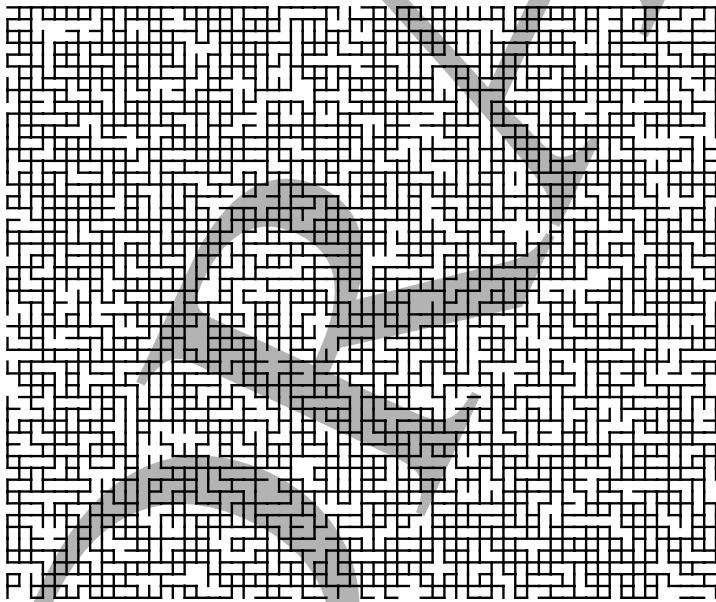
(a)  $p = 0.25$ (b)  $p = 0.49$ 

Figure 1.2. Realizations of bond percolation on a  $50 \times 60$  section of the square lattice for four different values of  $p$ . The pictures have been created using the same sequence of pseudorandom numbers, with the result that each graph is a subgraph of the next. Readers with good

(c)  $p = 0.51$ (d)  $p = 0.75$ 

eyesight may care to check that there exist open paths joining the left to the right side when  $p = 0.51$  but not when  $p = 0.49$ . The (random) value of  $p$  at which such paths appear for this realization is  $0.5059 \dots$ .

Thirdly, it is well endowed with beautiful conjectures which are easy to state but apparently rather hard to settle.

There is a fourth reason of significance. A great amount of effort has been invested in recent years towards an understanding of complex interacting random systems, including disordered media and other physical models. Such processes typically involve families of dependent random variables which are indexed by  $\mathbb{Z}^d$  for some  $d \geq 2$ . To develop a full theory of such a system is often beyond the current methodology. Instead, one may sometimes obtain partial results by making a comparison with another process which is better understood. It is often possible to make such a comparison with a percolation model. In this way, one may derive valuable results for the more complex system; these results may not be the best possible, but they may be compelling indicators of the directions to be pursued.

Here is an example. Consider a physical model having a parameter  $T$  called ‘temperature’. It may be suspected that there exists a critical value  $T_c$  marking a phase transition. While this fact may itself be unproven, it may be possible to prove by comparison that the behaviour of the process for small  $T$  is qualitatively different from that for large  $T$ .

It has been claimed that percolation theory is a cornerstone of the theory of disordered media. As evidence to support this claim, we make brief reference to four types of disordered physical systems, emphasizing the role of percolation for each.

**A. Disordered electrical networks.** It may not be too difficult to calculate the effective electrical resistance of a block of either material *A* or of material *B*, but what is the effective resistance of a mixture of these two materials? If the mixture is disordered, then it may be reasonable to assume that each component of the block is chosen at random to be of type *A* or of type *B*, independently of the types of all other components. The resulting effective resistance is a random variable whose distribution depends on the proportion  $p$  of components of type *A*. It seems to be difficult to say much of interest about the way in which this distribution depends on the numerical value of  $p$ . An extreme example arises when material *B* is a perfect insulator, and this is a case for which percolation comes to the fore. We illustrate this in a special example.

Let  $U_n$  be the square section  $\{0, 1, \dots, n\} \times \{0, 1, \dots, n\}$  of the square lattice, and let  $S_n$  and  $T_n$  be the bottom and top sides of  $U_n$ ,

$$S_n = \{(m, 0) : 0 \leq m \leq n\}, \quad T_n = \{(m, n) : 0 \leq m \leq n\}.$$

We turn  $U_n$  into an electrical network as follows. We examine each edge of  $U_n$  in turn, and replace it by a wire of resistance 1 Ohm with probability  $p$ , otherwise removing the connection entirely; this is done independently of all other edges. We now replace  $S_n$  and  $T_n$  by silver bars and we apply a potential difference between these bars; see Figure 1.3. What is the effective resistance  $R_n$  of the network?

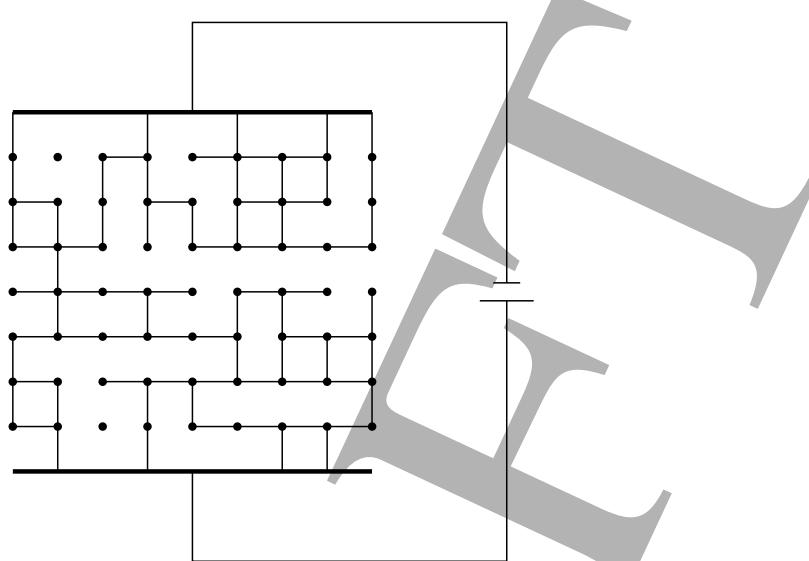


Figure 1.3. A realization of a random electrical network. Each remaining edge has unit resistance.

The value of  $R_n$  depends on the density and geometry of the set of edges having unit resistance, and such matters lie in the domain of percolation theory. We shall see in Section 13.2 that  $R_n = \infty$  for all large  $n$  (almost surely) if  $p < \frac{1}{2}$ , whereas  $R_n$  is bounded uniformly away from 0 and  $\infty$  if  $p > \frac{1}{2}$ .

**B. Ferromagnetism.** One of the most studied critical phenomena of theoretical physics is that of the ferromagnet. We position a lump of an appropriate metal in a magnetic field and we observe the way in which the magnetization of the metal varies according to imposed oscillations in the external field. Suppose that we increase the external field from 0 to some given value, and then decrease it back to 0. If the temperature is sufficiently large, the metal retains no residual magnetization, whereas at low temperatures the metal keeps some of its induced magnetization. There exists a critical value  $T_c$  of the temperature, called the Curie point, marking the borderline between the existence and non-existence of so called ‘spontaneous magnetization’. A standard mathematical model for this phenomenon is the ‘Ising model’. We give no definition of the Ising model here, but make instead some general remarks. In the Ising model on the lattice  $\mathbb{Z}^d$ , each vertex of  $\mathbb{Z}^d$  may be in either of two states labelled 0 and 1. A configuration is an assignment  $\omega = (\omega(x) : x \in \mathbb{Z}^d)$  of 0 or 1 to each vertex of the lattice. We consider probability measures on the set  $\Omega$  of configurations taken in conjunction with some suitable  $\sigma$ -field of subsets of  $\Omega$ ; in particular, we are concerned with a class of measures having a type of ‘spatial Markov’ property: conditional on the states of all vertices outside any finite connected subgraph  $G$  of  $\mathbb{Z}^d$ , the states of vertices of  $G$  depend only on those of vertices in its ‘external boundary’. Taken

in conjunction with certain other conditions (positivity and translation invariance of the conditional probabilities, and positive correlation of increasing events), this property characterizes the class of measures of interest for the model. It turns out that there are two parameters which specify the conditional probabilities: the ‘external magnetic field’  $h$ , and the strength  $J$  of interaction between neighbours. If  $J = 0$  then the states of different vertices are independent, and the process is equivalent to site percolation (see Section 1.6).

The relationship between the Ising model and bond percolation is rather strong. It turns out that they are linked via a type of ‘generalized percolation’ called the ‘random-cluster model’. Through studying the random-cluster model, one obtains conclusions valid simultaneously for percolation and the Ising model. This important discovery was made around 1970 by Fortuin and Kasteleyn, and it has greatly influenced part of the current view of disordered physical systems. See Section 13.6 for an account of the random-cluster model.

*C. Epidemics and fires in orchards.* In an early review of percolation and related topics, Frisch and Hammersley (1963) proposed the use of percolation in modelling the spread of blight in a large orchard. The problem is as follows. Hypothetical trees are grown at the vertices of a square lattice. We suppose that there is a probability  $p$  that a healthy tree will be infected by a neighbouring blighted tree, where  $p$  is a known function of the distance between neighbouring trees. To prevent a single blighted tree from endangering a significant proportion of the whole orchard, it is necessary to choose the lattice spacing to be large enough that  $p$  is smaller than the critical probability of bond percolation on  $\mathbb{Z}^2$ .

In a forest fire, trees which are completely destroyed by fire cannot threaten their neighbours. Similarly, trees which have recovered from measles presumably gain protection from recurrence of the disease. Such observations may be incorporated into a more complicated model which takes into account the passage of time. Suppose that each tree may be in any of three states: 1 (live and not on fire), 0 (burning), and  $-1$  (burned). We suppose that the tree at vertex  $x$  burns for a random time  $T_x$  after catching fire, where  $(T_x : x \in \mathbb{Z}^2)$  is a family of independent, identically distributed random variables. A burning tree emits sparks in the manner of a Poisson process with rate  $\alpha$ , and each spark hits one of the neighbouring trees chosen at random; the spark sets fire to that tree so long as it is neither burned nor already on fire. At time 0, an arsonist sets light to the tree at the origin. It turns out that the set  $C$  of trees which are ultimately burned in the ensuing conflagration may be identified as the set of vertices reachable from the origin by open paths of a certain percolation-type process; this process differs from ordinary bond percolation in that the states of two different edges may be dependent if the edges have a vertex in common. See Section 13.5, as well as Cox and Durrett (1988), van den Berg, Grimmett, and Schinazi (1998), and the references therein.

*D. Wafer-scale integration.* In the manufacture of microchips, silicon wafers are engraved with copies of the required circuitry, these copies being laid out

in a square grid. The wafer is then broken up into the individual chips, many of which are usually found to be faulty. After elimination of the faulty chips, the remaining non-defective chips are used to build processors. There are sound engineering and computing reasons for preferring to leave the wafer intact, making use of the non-defective chips in the positions in which they occurred in the wafer. Interconnections may be made between functioning units by using channels built between the rows and columns of the grid of chips. Such questions arise as the following:

- (i) How long is the longest linear chain of functioning units which may be created using interconnections each of length not exceeding  $\delta$  lattice units and laid in such a way that each channel of the wafer contains no more than two such interconnections?
- (ii) Find the minimal interconnection length in a wiring pattern which creates a square grid of size  $k \times k$  of functioning chips out of a wafer containing  $n \times n$  units in all.

Greene and El Gamal (1984) answer such questions under the hypothesis that each chip is non-defective with probability  $p$ , independently of all other chips. Under this assumption, the set of functioning chips may be identified as the set of open vertices in a site percolation process on  $\mathbb{Z}^2$ , and thus the theory of percolation is important.

### 1.3 Bond Percolation

In this section we shall establish the basic definitions and notation of bond percolation on  $\mathbb{Z}^d$ . We begin with some graph theory. Throughout most of this book, the letter  $d$  stands for the dimension of the process; generally  $d \geq 2$ , but we assume for the moment only that  $d \geq 1$ . We write  $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$  for the set of all integers, and  $\mathbb{Z}^d$  for the set of all vectors  $x = (x_1, x_2, \dots, x_d)$  with integral coordinates. For  $x \in \mathbb{Z}^d$  we generally write  $x_i$  for the  $i$ th coordinate of  $x$ . The (graph-theoretic) distance  $\delta(x, y)$  from  $x$  to  $y$  is defined by

$$(1.1) \quad \delta(x, y) = \sum_{i=1}^d |x_i - y_i|,$$

and we write  $|x|$  for the distance  $\delta(0, x)$  from the origin to  $x$ . We shall sometimes have use for another distance function on  $\mathbb{Z}^d$ , and shall write

$$(1.2) \quad \|x\| = \max\{|x_i| : 1 \leq i \leq d\},$$

noting that

$$\|x\| \leq |x| \leq d\|x\|.$$

We may turn  $\mathbb{Z}^d$  into a graph, called the *d-dimensional cubic lattice*, by adding edges between all pairs  $x, y$  of points of  $\mathbb{Z}^d$  with  $\delta(x, y) = 1$ . We denote this

lattice by  $\mathbb{L}^d$ , and we write  $\mathbb{Z}^d$  for the set of vertices of  $\mathbb{L}^d$ , and  $\mathbb{E}^d$  for the set of its edges. In graph-theoretic terms, we write  $\mathbb{L}^d = (\mathbb{Z}^d, \mathbb{E}^d)$ . We shall often think of  $\mathbb{L}^d$  as a graph embedded in  $\mathbb{R}^d$ , the edges being straight line segments between their endvertices. If  $\delta(x, y) = 1$ , then we say that  $x$  and  $y$  are *adjacent*; in this case, we write  $x \sim y$  and we represent the edge from  $x$  to  $y$  as  $\langle x, y \rangle$ . The edge  $e$  is *incident* to the vertex  $x$  if  $x$  is an endvertex of  $e$ . Letters such as  $u, v, w, x, y$  usually represent vertices, and letters such as  $e, f$  usually represent edges. We denote the origin of  $\mathbb{Z}^d$  by 0.

Next we introduce probability. Let  $p$  and  $q$  satisfy  $0 \leq p \leq 1$  and  $p + q = 1$ . We declare each edge of  $\mathbb{L}^d$  to be *open* with probability  $p$  and *closed* otherwise, independently of all other edges. More formally, we consider the following probability space. As sample space we take  $\Omega = \prod_{e \in \mathbb{E}^d} \{0, 1\}$ , points of which are represented as  $\omega = (\omega(e) : e \in \mathbb{E}^d)$  and called *configurations*; the value  $\omega(e) = 0$  corresponds to  $e$  being closed, and  $\omega(e) = 1$  corresponds to  $e$  being open. We take  $\mathcal{F}$  to be the  $\sigma$ -field of subsets of  $\Omega$  generated by the finite-dimensional cylinders. Finally, we take product measure with density  $p$  on  $(\Omega, \mathcal{F})$ ; this is the measure

$$P_p = \prod_{e \in \mathbb{E}^d} \mu_e$$

where  $\mu_e$  is Bernoulli measure on  $\{0, 1\}$ , given by

$$\mu_e(\omega(e) = 0) = q, \quad \mu_e(\omega(e) = 1) = p.$$

We write  $P_p$  for this product measure, and  $E_p$  for the corresponding expectation operator. We shall occasionally need a more general construction in which different edges may have different probabilities of being open. Such a construction begins with a family  $\mathbf{p} = (p(e) : e \in \mathbb{E}^d)$  with  $0 \leq p(e) \leq 1$  for all  $e$ . The appropriate probability space is now  $(\Omega, \mathcal{F}, P_{\mathbf{p}})$  where  $P_{\mathbf{p}} = \prod_{e \in \mathbb{E}^d} \mu_e$  and

$$\mu_e(\omega(e) = 0) = 1 - p(e), \quad \mu_e(\omega(e) = 1) = p(e)$$

for each  $e$ .

We write  $\overline{A}$  (or occasionally  $A^c$ ) for the complement of an event  $A$ , and  $I_A$  for the indicator function of  $A$ :

$$I_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{if } \omega \notin A. \end{cases}$$

The expression  $E_p(X; A)$  denotes the mean of  $X$  on the event  $A$ ; that is to say,  $E_p(X; A) = E_p(X I_A)$ .

The following notation will be of value later. Let  $f$  be an edge of  $\mathbb{L}^d$ . We write  $P_p^f$  for Bernoulli product measure on  $\prod_{e: e \neq f} \{0, 1\}$ , the set of configurations of all edges of the lattice other than  $f$ . We think of  $P_p^f$  as being the measure associated with percolation on  $\mathbb{L}^d$  with the edge  $f$  deleted.

There is a natural partial order on the set  $\Omega$  of configurations, given by  $\omega_1 \leq \omega_2$  if and only if  $\omega_1(e) \leq \omega_2(e)$  for all  $e \in \mathbb{E}^d$ .

There is a one-one correspondence between  $\Omega$  and the set of subsets of  $\mathbb{E}^d$ . For  $\omega \in \Omega$ , we define

$$(1.3) \quad K(\omega) = \{e \in \mathbb{E}^d : \omega(e) = 1\};$$

thus  $K(\omega)$  is the set of open edges of the lattice when the configuration is  $\omega$ . Clearly,  $\omega_1 \leq \omega_2$  if and only if  $K(\omega_1) \subseteq K(\omega_2)$ .

The following device can be useful. Suppose that  $(X(e) : e \in \mathbb{E}^d)$  is a family of independent random variables indexed by the edge set  $\mathbb{E}^d$ , where each  $X(e)$  is uniformly distributed on  $[0, 1]$ . We may couple together *all* bond percolation processes on  $\mathbb{L}^d$  as  $p$  ranges over the interval  $[0, 1]$  in the following way. Let  $p$  satisfy  $0 \leq p \leq 1$  and define  $\eta_p$  ( $\in \Omega$ ) by

$$(1.4) \quad \eta_p(e) = \begin{cases} 1 & \text{if } X(e) < p, \\ 0 & \text{if } X(e) \geq p. \end{cases}$$

We say that the edge  $e$  is  $p$ -open if  $\eta_p(e) = 1$ . The random vector  $\eta_p$  has independent components and marginal distributions given by

$$P(\eta_p(e) = 0) = 1 - p, \quad P(\eta_p(e) = 1) = p.$$

We may think of  $\eta_p$  as being the random outcome of the bond percolation process on  $\mathbb{L}^d$  with edge-probability  $p$ . It is clear that  $\eta_{p_1} \leq \eta_{p_2}$  whenever  $p_1 \leq p_2$ , which is to say that we may couple together the two percolation processes with edge-probabilities  $p_1$  and  $p_2$  in such a way that the set of open edges of the first process is a subset of the set of open edges of the second. More generally, as  $p$  increases from 0 to 1, the configuration  $\eta_p$  runs through typical configurations of percolation processes with all edge-probabilities.

A path of  $\mathbb{L}^d$  is an alternating sequence  $x_0, e_0, x_1, e_1, \dots, e_{n-1}, x_n$  of distinct vertices  $x_i$  and edges  $e_i = \langle x_i, x_{i+1} \rangle$ ; such a path has length  $n$  and is said to connect  $x_0$  to  $x_n$ . A circuit of  $\mathbb{L}^d$  is an alternating sequence  $x_0, e_0, x_1, e_1, \dots, e_{n-1}, x_n, e_n$ ,  $x_0$  of vertices and edges such that  $x_0, e_0, \dots, e_{n-1}, x_n$  is a path and  $e_n = \langle x_n, x_0 \rangle$ ; such a circuit has length  $n + 1$ . We call a path or circuit open if all of its edges are open, and closed if all of its edges are closed. Two subgraphs of  $\mathbb{L}^d$  are called edge-disjoint if they have no edges in common, and disjoint if they have neither edges nor vertices in common.

Consider the random subgraph of  $\mathbb{L}^d$  containing the vertex set  $\mathbb{Z}^d$  and the open edges only. The connected components of this graph are called *open clusters*. We write  $C(x)$  for the open cluster containing the vertex  $x$ , and we call  $C(x)$  the *open cluster at  $x$* . The vertex set of  $C(x)$  is the set of all vertices of the lattice which are connected to  $x$  by open paths, and the edges of  $C(x)$  are the open edges of  $\mathbb{L}^d$  which join pairs of such vertices. By the translation invariance of the lattice and of the probability measure  $P_p$ , the distribution of  $C(x)$  is independent of the choice

of  $x$ . The open cluster  $C(0)$  at the origin is therefore typical of such clusters, and we represent this cluster by the single letter  $C$ . Occasionally we shall use the term  $C(x)$  to represent the *set of vertices* joined to  $x$  by open paths, rather than the graph of this open cluster. We shall be interested in the size of  $C(x)$ , and we denote by  $|C(x)|$  the number of vertices in  $C(x)$ . We note that  $C(x) = \{x\}$  whenever  $x$  is incident to no open edge.

If  $A$  and  $B$  are sets of vertices of  $\mathbb{L}^d$ , we shall write ‘ $A \leftrightarrow B$ ’ if there exists an open path joining some vertex in  $A$  to some vertex in  $B$ ; if  $A \cap B \neq \emptyset$  then  $A \leftrightarrow B$  trivially. Thus, for example,  $C(x) = \{y \in \mathbb{Z}^d : x \leftrightarrow y\}$ . We shall write ‘ $A \not\leftrightarrow B$ ’ if there exists no open path from any vertex of  $A$  to any vertex of  $B$ , and ‘ $A \leftrightarrow B$  off  $D$ ’ if there exists an open path joining some vertex in  $A$  to some vertex in  $B$  which uses no vertex in the set  $D$ .

If  $A$  is a set of vertices of the lattice, we write  $\partial A$  for the *surface* of  $A$ , being the set of vertices in  $A$  which are adjacent to some vertex not in  $A$ .

Our notation for boxes is the following. A *box* is a subset of  $\mathbb{Z}^d$  of the form  $B(a, b) = \{x \in \mathbb{Z}^d : a_i \leq x_i \leq b_i \text{ for all } i\}$ , where  $a$  and  $b$  lie in  $\mathbb{Z}^d$ ; we sometimes write

$$B(a, b) = \prod_{i=1}^d [a_i, b_i]$$

as a convenient shorthand. We often think of  $B(a, b)$  as a subgraph of the lattice  $\mathbb{L}^d$  suitably endowed with the edges which it inherits from the lattice. We denote by  $B(n)$  the box with side-length  $2n$  and centre at the origin:

$$(1.5) \quad B(n) = [-n, n]^d = \{x \in \mathbb{Z}^d : \|x\| \leq n\}.$$

We may turn  $B(n)$  into a graph by adding the edges which it inherits from  $\mathbb{L}^d$ . If  $x$  is a vertex of the lattice, we write  $B(n, x)$  for the box  $x + B(n)$  having side-length  $2n$  and centre at  $x$ .

We write  $\lfloor a \rfloor$  and  $\lceil a \rceil$  for the integer part of the real number  $a$ , and the least integer not less than  $a$ , respectively. If  $(a_n : n \geq 1)$  and  $(b_n : n \geq 1)$  are sequences of real numbers, we write  $a_n \sim b_n$  if  $a_n/b_n \rightarrow 1$  as  $n \rightarrow \infty$ , and  $a_n \approx b_n$  if  $\log a_n / \log b_n \rightarrow 1$  as  $n \rightarrow \infty$ . Similarly, we write  $f(p) \sim g(p)$  (respectively  $f(p) \approx g(p)$ ) as  $p \rightarrow \pi$  if  $f(p)/g(p) \rightarrow 1$  (respectively  $\log f(p)/\log g(p) \rightarrow 1$ ) as  $p \rightarrow \pi$ . Finally, we write  $f(p) \asymp g(p)$  as  $p \rightarrow \pi$  if  $f(p)/g(p)$  is bounded away from 0 and  $\infty$  on a neighbourhood of  $\pi$ .

## 1.4 The Critical Phenomenon

A principal quantity of interest is the *percolation probability*  $\theta(p)$ , being the probability that a given vertex belongs to an infinite open cluster. By the translation invariance of the lattice and probability measure, we lose no generality by taking

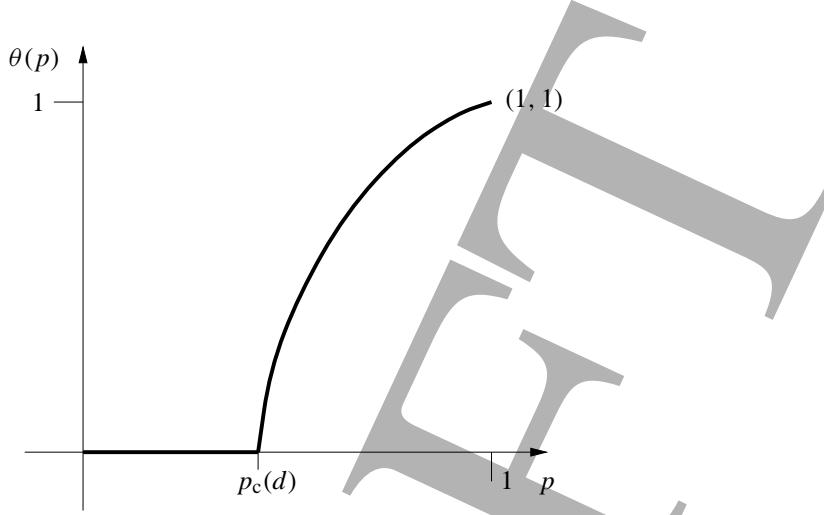


Figure 1.4. It is believed that the percolation probability  $\theta(p)$  behaves roughly as indicated. It is known, for example, that  $\theta$  is a continuous function of  $p$  except possibly at the critical probability  $p_c(d)$ . The possibility of a jump discontinuity at  $p_c(d)$  has not been ruled out when  $3 \leq d < 19$ .

this vertex to be the origin, and thus we define

$$(1.6) \quad \theta(p) = P_p(|C| = \infty).$$

Alternatively, we may write

$$(1.7) \quad \theta(p) = 1 - \sum_{n=1}^{\infty} P_p(|C| = n).$$

It is easy to see that  $|C| = \infty$  if and only if there exists an infinite sequence  $x_0, x_1, x_2, \dots$  of distinct vertices such that  $x_0 = 0$ ,  $x_i \sim x_{i+1}$ , and  $\langle x_i, x_{i+1} \rangle$  is open for all  $i$ . Clearly  $\theta$  is a non-decreasing function of  $p$  with  $\theta(0) = 0$  and  $\theta(1) = 1$ . (Probably the most transparent proof of this monotonicity makes use of the coupling introduced around (1.4). See also Section 2.1.)

It is fundamental to percolation theory that there exists a critical value  $p_c = p_c(d)$  of  $p$  such that

$$\theta(p) \begin{cases} = 0 & \text{if } p < p_c, \\ > 0 & \text{if } p > p_c; \end{cases}$$

$p_c(d)$  is called the *critical probability* and is defined formally by

$$(1.8) \quad p_c(d) = \sup\{p : \theta(p) = 0\}.$$

The case of one dimension is of no interest since, if  $p < 1$ , there exist infinitely many closed edges of  $\mathbb{L}^1$  to the left and to the right of the origin almost surely,

implying that  $\theta(p) = 0$  if  $p < 1$ ; thus  $p_c(1) = 1$ . The situation is quite different in two and more dimensions, and we shall see in Theorem (1.10) that  $0 < p_c(d) < 1$  if  $d \geq 2$ . We shall assume henceforth that, in the absence of an indication to the contrary,  $d$  is at least 2. See Figure 1.4 for a sketch of the function  $\theta$ .

The  $d$ -dimensional lattice  $\mathbb{L}^d$  may be embedded in  $\mathbb{L}^{d+1}$  in a natural way as the projection of  $\mathbb{L}^{d+1}$  onto the subspace generated by the first  $d$  coordinates; with this embedding, the origin of  $\mathbb{L}^{d+1}$  belongs to an infinite open cluster for a particular value of  $p$  whenever it belongs to an infinite open cluster of the sublattice  $\mathbb{L}^d$ . Thus  $\theta(p) = \theta_d(p)$  is non-decreasing in  $d$ , which implies that

$$(1.9) \quad p_c(d+1) \leq p_c(d) \quad \text{for } d \geq 1.$$

It is not very difficult to show that strict inequality is valid here, in that  $p_c(d+1) < p_c(d)$  for all  $d \geq 1$ ; see Sections 1.7 and 3.3.

The following theorem amounts to the statement that there exists a non-trivial critical phenomenon in dimensions two and more.

**(1.10) Theorem.** *If  $d \geq 2$  then  $0 < p_c(d) < 1$ .*

The nub of this theorem is that in two or more dimensions there are two phases of the process. In the *subcritical* phase when  $p < p_c(d)$ , every vertex is almost surely in a finite open cluster, so that all open clusters are almost surely finite. In the *supercritical* phase when  $p > p_c(d)$ , each vertex has a strictly positive probability of being in an infinite open cluster, so that there exists almost surely at least one infinite open cluster. These phases are now reasonably well understood, which is more than can be said about the intermediate *critical* percolation process with  $p = p_c(d)$ , to which we shall return in more detail in Chapter 9. We make more concrete the above remarks about the subcritical and supercritical phases.

**(1.11) Theorem.** *The probability  $\psi(p)$  that there exists an infinite open cluster satisfies*

$$\psi(p) = \begin{cases} 0 & \text{if } p < p_c(d), \\ 1 & \text{if } p > p_c(d). \end{cases}$$

This theorem says nothing about the existence or non-existence of infinite open clusters when  $p = p_c(d)$ . It turns out that no infinite open cluster exists when either  $d = 2$  or  $d \geq 19$ , but it is an open question to determine whether or not there exists such a cluster for general  $d$  (including the physically important case  $d = 3$ ); it is expected that no such cluster exists. Theorem (1.11) is proved by an application of the zero–one law, and this tells us nothing about the actual *number* of infinite open clusters when  $p > p_c(d)$ ; we shall however see in Section 8.2 that the infinite open cluster is (almost surely) unique whenever it exists.

Before proving these two theorems, we mention some associated results and open problems. First, what is the numerical value of  $p_c(d)$ ? We know only the values  $p_c(1) = 1$  and  $p_c(2) = \frac{1}{2}$ . The latter value is far from trivial to show, and this was the prize which attracted many people to the field in the 1970s. It

is highly unlikely that there exists a useful representation of  $p_c(d)$  for any other value of  $d$ , although such values may be calculated with increasing degrees of accuracy with the aid of larger and faster computers. Exact values are known for the critical probabilities of certain other two-dimensional lattices (for example,  $p_c = 2 \sin(\pi/18)$  for bond percolation on the triangular lattice); see Sections 3.1 and 11.9. It is the case that the value of the critical probability depends on both the dimension and the lattice structure, in contrast to certain asymptotic properties of  $\theta(p)$  when  $p$  is near  $p_c$ : it is thought that, when  $p - p_c$  is small and positive, then  $\theta(p)$  behaves in a way which depends, to a degree, on the dimension  $d$  alone and is independent of the particular lattice structure. We return to this point in the next section and in Chapter 9.

Secondly, it is not difficult to find non-trivial upper and lower bounds for  $p_c(d)$  when  $d \geq 2$ . We shall see in the proof of Theorem (1.10) that

$$(1.12) \quad \frac{1}{\lambda(2)} \leq p_c(2) \leq 1 - \frac{1}{\lambda(2)},$$

and more generally

$$(1.13) \quad \frac{1}{\lambda(d)} \leq p_c(d) \quad \text{for } d \geq 3;$$

here,  $\lambda(d)$  is the *connective constant* of  $\mathbb{L}^d$ , given by

$$(1.14) \quad \lambda(d) = \lim_{n \rightarrow \infty} \{\sigma(n)^{1/n}\},$$

where  $\sigma(n)$  is the number of paths (or ‘self-avoiding walks’) of  $\mathbb{L}^d$  having length  $n$  and beginning at the origin. The exact value of  $\lambda(d)$  is unknown for  $d \geq 2$ , but it is obvious that  $\lambda(d) \leq 2d - 1$ ; to see this, note that each new step in a self-avoiding walk has at most  $2d - 1$  choices since it must avoid the current position, and therefore  $\sigma(n) \leq 2d(2d - 1)^{n-1}$ .

Thirdly, how does  $p_c(d)$  behave when  $d$  is large? Inequality (1.13) implies that  $(2d - 1)p_c(d) \geq 1$ , and it is known further that  $p_c(d) \sim (2d)^{-1}$  as  $d \rightarrow \infty$ . This amounts to saying that, for large  $d$ , bond percolation on  $\mathbb{L}^d$  behaves similarly to bond percolation on a regular tree in which each vertex has  $2d(1 + o(1))$  descendants.

**Proof of Theorem (1.10) and Equation (1.12).** The existence of a critical phenomenon was shown by Broadbent and Hammersley (1957) and Hammersley (1957a, 1959).

We saw in (1.9) that  $p_c(d+1) \leq p_c(d)$ , and it suffices therefore to show that  $p_c(d) > 0$  for  $d \geq 2$ , and that  $p_c(2) < 1$ . We prove first that  $p_c(d) > 0$  for  $d \geq 2$ . Consider bond percolation on  $\mathbb{L}^d$  when  $d \geq 2$ . We shall show that  $\theta(p) = 0$  whenever  $p$  is sufficiently close to 0. Let  $\sigma(n)$  be the number of paths of

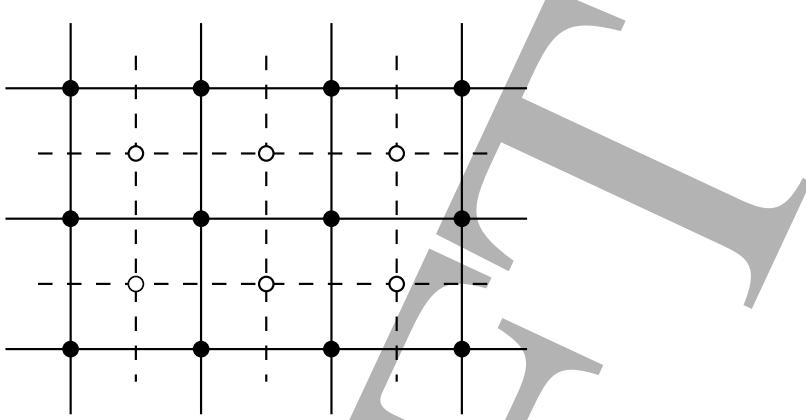


Figure 1.5. Part of the square lattice  $\mathbb{L}^2$  together with its dual.

$\mathbb{L}^d$  which have length  $n$  and which begin at the origin, and let  $N(n)$  be the number of such paths which are open. Any such path is open with probability  $p^n$ , so that

$$E_p(N(n)) = p^n \sigma(n).$$

Now, if the origin belongs to an infinite open cluster then there exist open paths of all lengths beginning at the origin, so that

$$(1.15) \quad \begin{aligned} \theta(p) &\leq P_p(N(n) \geq 1) \\ &\leq E_p(N(n)) = p^n \sigma(n) \quad \text{for all } n. \end{aligned}$$

By the definition of the connective constant  $\lambda(d)$  given at (1.14), we have that  $\sigma(n) = \{\lambda(d) + o(1)\}^n$  as  $n \rightarrow \infty$ ; we substitute this into (1.15) to obtain

$$(1.16) \quad \begin{aligned} \theta(p) &\leq \{p\lambda(d) + o(1)\}^n \\ &\rightarrow 0 \quad \text{if } p\lambda(d) < 1 \end{aligned}$$

as  $n \rightarrow \infty$ . Thus we have shown that  $p_c(d) \geq \lambda(d)^{-1}$  where  $\lambda(d) \leq 2d - 1 < \infty$ .

Secondly, we show that  $p_c(2) < 1$ , and we use an approach which is commonly called a ‘Peierls argument’ in honour of Rudolf Peierls and his 1936 article on the Ising model. Consider bond percolation on  $\mathbb{L}^2$ ; we shall show that  $\theta(p) > 0$  if  $p$  is sufficiently close to 1. It is here that planar duality is useful. Let  $G$  be a planar graph, drawn in the plane in such a way that edges intersect at vertices only. The *planar dual* of  $G$  is the graph obtained from  $G$  in the following way. We place a vertex in each face of  $G$  (including any infinite faces which may exist) and join two such vertices by an edge whenever the corresponding faces of  $G$  share a boundary edge in  $G$ . It is easy to see (especially with the aid of Figure 1.5) that the dual of  $\mathbb{L}^2$  is isomorphic to  $\mathbb{L}^2$ ; this *self-duality* is not in itself important at this stage, but will be crucial to our forthcoming proof in Chapter 11 that  $p_c(2) = \frac{1}{2}$ . For the sake of

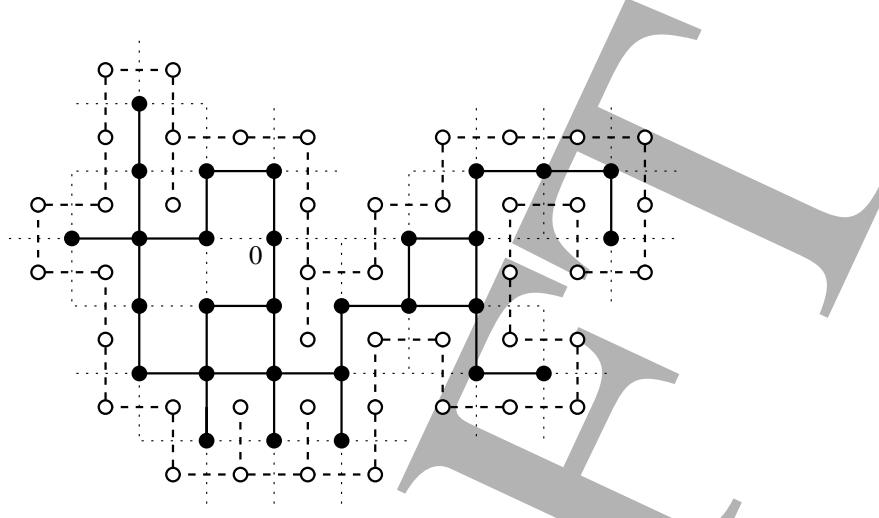


Figure 1.6. A finite open cluster at the origin, surrounded by a closed circuit in the dual lattice.

definiteness, we take as vertices of this dual lattice the set  $\{x + (\frac{1}{2}, \frac{1}{2}) : x \in \mathbb{Z}^2\}$  and we join two such neighbouring vertices by a straight line segment of  $\mathbb{R}^2$ . There is a one-one correspondence between the edges of  $\mathbb{L}^2$  and the edges of the dual, since each edge of  $\mathbb{L}^2$  is crossed by a unique edge of the dual. We declare an edge of the dual to be open or closed depending respectively on whether it crosses an open or closed edge of  $\mathbb{L}^2$ . This assignment gives rise to a bond percolation process on the dual lattice with the same edge-probability  $p$ . We shall return to such matters in Chapter 11.

Suppose now that the open cluster at the origin of  $\mathbb{L}^2$  is finite, and see Figure 1.6 for a sketch of the situation. We see that the origin is surrounded by a necklace of closed edges which are blocking off all possible routes from the origin to infinity. We may satisfy ourselves that the corresponding edges of the dual contain a closed circuit in the dual having the origin of  $\mathbb{L}^2$  in its interior. This is best seen by inspecting Figure 1.6 again. It is somewhat tedious to formulate and prove such a statement with complete rigour, and we shall not do so here; see Kesten (1982, p. 386) for a more careful treatment. The converse holds similarly: if the origin lies in the interior of a closed circuit of the dual lattice, then the open cluster at the origin is finite. We summarize these remarks by saying that  $|C| < \infty$  if and only if the origin of  $\mathbb{L}^2$  lies in the interior of some closed circuit of the dual.

We now proceed as in the first part of the proof, by counting the number of such closed circuits in the dual. Let  $\rho(n)$  be the number of circuits in the dual which have length  $n$  and which contain in their interiors the origin of  $\mathbb{L}^2$ . We may estimate  $\rho(n)$  as follows. Each such circuit passes through some vertex of the form  $(k + \frac{1}{2}, \frac{1}{2})$  for some  $k$  satisfying  $0 \leq k < n$  since: first, it surrounds the origin and therefore passes through  $(k + \frac{1}{2}, \frac{1}{2})$  for some  $k \geq 0$  and, secondly, it

cannot pass through  $(k + \frac{1}{2}, \frac{1}{2})$  where  $k \geq n$  since then it would have length at least  $2n$ . Thus such a circuit contains a self-avoiding walk of length  $n - 1$  starting from a vertex of the form  $(k + \frac{1}{2}, \frac{1}{2})$  where  $0 \leq k < n$ . The number of such self-avoiding walks is at most  $n\sigma(n - 1)$ , giving that

$$(1.17) \quad \rho(n) \leq n\sigma(n - 1).$$

Let  $\gamma$  be a circuit of the dual containing the origin of  $\mathbb{L}^2$  in its interior, and let  $M(n)$  be the number of such closed circuits having length  $n$ . By (1.17),

$$\begin{aligned} (1.18) \quad \sum_{\gamma} P_p(\gamma \text{ is closed}) &\leq \sum_{n=1}^{\infty} q^n n\sigma(n - 1) \\ &= \sum_{n=1}^{\infty} qn \{q\lambda(2) + o(1)\}^{n-1} \quad \text{as in (1.16)} \\ &< \infty \quad \text{if } q\lambda(2) < 1, \end{aligned}$$

where  $q = 1 - p$  and the summation is over all such  $\gamma$ . Furthermore,

$$\sum_{\gamma} P_p(\gamma \text{ is closed}) \rightarrow 0 \quad \text{as } q = 1 - p \downarrow 0,$$

so that we may find  $\pi$  satisfying  $0 < \pi < 1$  such that

$$\sum_{\gamma} P_p(\gamma \text{ is closed}) \leq \frac{1}{2} \quad \text{if } p > \pi.$$

It follows from the previous remarks that

$$\begin{aligned} P_p(|C| = \infty) &= P_p(M(n) = 0 \text{ for all } n) \\ &= 1 - P_p(M(n) \geq 1 \text{ for some } n) \\ &\geq 1 - \sum_{\gamma} P_p(\gamma \text{ is closed}) \\ &\geq \frac{1}{2} \quad \text{if } p > \pi, \end{aligned}$$

giving that  $p_c(2) \leq \pi$ .

We need to work slightly harder in order to deduce that  $p_c(2) \leq 1 - \lambda(2)^{-1}$ , as required for (1.12). The usual proof of this makes use of a certain correlation inequality known as the FKG inequality, which will be presented in Section 2.2. Rather than follow this usual route, we use a more elementary method which requires no extra technology. Let  $m$  be a positive integer. Let  $F_m$  be the event that there exists a closed dual circuit containing the box  $B(m)$  in its interior, and let  $G_m$

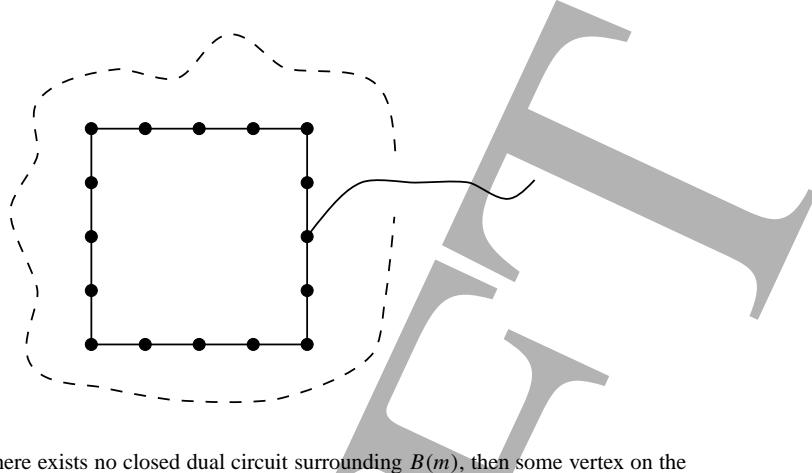


Figure 1.7. If there exists no closed dual circuit surrounding  $B(m)$ , then some vertex on the surface of  $B(m)$  lies in an infinite open path.

be the event that all edges of  $B(m)$  are open. These two events are independent, since they are defined in terms of disjoint sets of edges. Now, similarly to (1.18),

$$P_p(F_m) \leq P_p\left(\sum_{n=4m}^{\infty} M(n) \geq 1\right) \leq \sum_{n=4m}^{\infty} q^n n \sigma(n-1).$$

Much as before, if  $q < \lambda(2)^{-1}$ , we may find  $m$  such that  $P_p(F_m) < \frac{1}{2}$ , and we choose  $m$  accordingly. Assume now that  $G_m$  occurs but  $F_m$  does not. As indicated in Figure 1.7, the non-occurrence of  $F_m$  implies that some vertex of  $B(m)$  lies in an infinite open path. Combined with the occurrence of  $G_m$ , this implies that  $|C| = \infty$ . Therefore, using the independence of  $F_m$  and  $G_m$ ,

$$\theta(p) \geq P_p(\overline{F_m} \cap G_m) = P_p(\overline{F_m}) P_p(G_m) \geq \frac{1}{2} P_p(G_m) > 0$$

if  $q < \lambda(2)^{-1}$ . □

**Proof of Theorem (1.11).** This is straightforward. First, we note that the event  $\{\mathbb{L}^d \text{ contains an infinite open cluster}\}$  does not depend upon the states of any finite collection of edges. By the usual zero–one law (see, for example, Grimmett and Stirzaker (1992, p. 290)),  $\psi$  takes the values 0 and 1 only. If  $\theta(p) = 0$  then

$$\psi(p) \leq \sum_{x \in \mathbb{Z}^d} P_p(|C(x)| = \infty) = 0.$$

On the other hand, if  $\theta(p) > 0$  then

$$\psi(p) \geq P_p(|C| = \infty) > 0$$

so that  $\psi(p) = 1$  by the zero–one law, as required. □

## 1.5 The Main Questions

Consider bond percolation on  $\mathbb{L}^d$  where  $d \geq 2$ . We are interested in the sizes and shapes of typical open clusters as the edge-probability  $p$  varies from 0 to 1, and we are particularly interested in large-scale phenomena such as the existence of infinite open clusters. We saw in the last section that ‘macroscopic’ quantities such as  $\theta(p)$  and  $\chi(p)$  have qualitatively different behaviour for small  $p$  than they have for large  $p$ . In addition to the probability that an open cluster is infinite, we may be interested in the mean size of an open cluster, and we write

$$(1.19) \quad \chi(p) = E_p|C|$$

for the mean number of vertices in the open cluster at the origin. By the translation invariance of the process, we have that  $\chi(p) = E_p|C(x)|$  for all vertices  $x$ . The functions  $\theta$  and  $\chi$  are two of the principal characters in percolation theory. We may express  $\chi$  in terms of the distribution of  $|C|$ , just as we did for  $\theta$  in (1.7):

$$(1.20) \quad \begin{aligned} \chi(p) &= \infty \cdot P_p(|C| = \infty) + \sum_{n=1}^{\infty} n P_p(|C| = n) \\ &= \infty \cdot \theta(p) + \sum_{n=1}^{\infty} n P_p(|C| = n), \end{aligned}$$

so that

$$(1.21) \quad \chi(p) = \infty \quad \text{if } p > p_c.$$

The converse is not at all obvious: is it the case that  $\chi(p) < \infty$  if  $p < p_c$ ? We answer this question affirmatively in Chapter 5 (a sketch of the function  $\chi$  appears in Figure 1.8). This indicates that the ‘macroscopic’ quantities  $\theta$  and  $\chi$  manifest critical behaviour at the same value of  $p$ . Indeed, most ‘reasonable’ macroscopic functions, such as  $\theta$  and  $\chi$ , are smooth functions of  $p$  except at the critical value  $p_c$ . It is commonly said that there exists a unique phase transition for percolation. More precisely, there are exactly two phases in the model—the subcritical phase ( $p < p_c$ ) and the supercritical phase ( $p > p_c$ )—together with the process at the critical point (when  $p = p_c$ ). We shall study these phases in some detail in Chapters 5–10, but we present here a brief preview of some of the main results and open problems.

*Subcritical phase.* When  $p < p_c$ , all open clusters are finite almost surely. We shall see in Chapter 6 that  $|C|$  has a tail which decreases exponentially, which is to say that there exists  $\alpha(p)$  such that

$$(1.22) \quad P_p(|C| = n) \approx e^{-n\alpha(p)} \quad \text{as } n \rightarrow \infty,$$

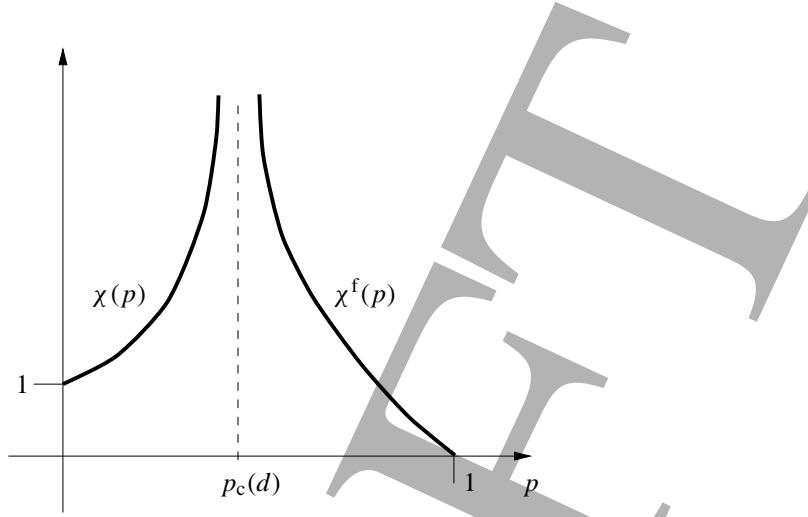


Figure 1.8. The left-hand curve is a sketch of the mean cluster size  $\chi(p)$ . The right-hand curve is a sketch of the mean size  $\chi^f(p)$  of a finite open cluster when  $p > p_c$ . Note that  $\chi(p) = \chi^f(p)$  if  $p < p_c(d)$ .

and  $\alpha(p) > 0$  when  $p < p_c$ . It follows that  $|C|$  has finite moments of all orders when  $p < p_c$ .

*Supercritical phase.* When  $p > p_c$ , there exist infinite open clusters almost surely, but how many? We shall see in Section 8.2 that the infinite open cluster is unique almost surely. If  $|C| < \infty$  then how fast does the tail of  $|C|$  decay? It is known that there exist  $\beta_1(p)$  and  $\beta_2(p)$ , satisfying  $0 < \beta_2(p) \leq \beta_1(p) < \infty$ , such that

$$(1.23) \quad \exp(-\beta_1(p)n^{(d-1)/d}) \leq P_p(|C| = n) \leq \exp(-\beta_2(p)n^{(d-1)/d}) \quad \text{for all } n,$$

and it is believed that the limit

$$(1.24) \quad \delta(p) = \lim_{n \rightarrow \infty} \left\{ -n^{-(d-1)/d} \log P_p(|C| = n) \right\}$$

exists and is strictly positive when  $p > p_c$ . The basic reason for the power  $n^{(d-1)/d}$  is that this is the order of the surface area of the sphere in  $\mathbb{R}^d$  with volume  $n$ . The existence of the limit in (1.24) has been proved when  $d = 2$  by Alexander, Chayes, and Chayes (1990), and when  $d = 3$  by Cerf (1998b).

Since  $\chi(p) = \infty$  when  $p > p_c$ , the function  $\chi$  is of little interest in the supercritical phase. Instead, we concentrate on the related ‘truncated’ function given by

$$(1.25) \quad \chi^f(p) = E_p(|C|; |C| < \infty),$$

the mean of  $|C|$  on the event that  $|C| < \infty$ . The function  $\chi^f$  is probably not dissimilar in general form to the sketch in Figure 1.8. The superscript ‘f’ refers to the condition that  $C$  be finite.

*At the critical point.* It is hereabouts that we find major open problems. First, does there exist an infinite open cluster when  $p = p_c$ ? The answer is known to be negative when  $d = 2$  or  $d \geq 19$ , and is generally believed to be negative for all  $d \geq 2$ . Assuming that  $\theta(p_c) = 0$ , which is to say that there exists no infinite open cluster when  $p = p_c$ , at what rate does  $P_{p_c}(|C| = n)$  decay? It is believed that

$$(1.26) \quad P_{p_c}(|C| \geq n) \approx n^{-1/\delta} \quad \text{as } n \rightarrow \infty$$

for some  $\delta = \delta(d) > 0$ ; the quantity  $\delta$  is an example of a ‘critical exponent’. Lower bounds for  $P_p(|C| \geq n)$  of this general ‘power’ form are known for all dimensions  $d \geq 2$ , and also upper bounds when  $d = 2$ . Some have asked the provocative question “is it true that  $\delta = \frac{91}{5}$  when  $d = 2$ , and  $\delta = 2$  when  $d \geq 6$ ?”; see Newman (1987a), for example.

Major progress has been made towards an understanding of critical percolation, but only under the assumption that  $d$  is sufficiently large. Currently the condition  $d \geq 19$  suffices. When this holds, we know that  $\theta(p_c) = 0$ , together with exact calculations of certain critical exponents.

*Near the critical point.* As  $p$  approaches  $p_c$  from above (or beneath), how do such quantities as  $\theta(p)$  (or  $\chi(p)$ ) behave? It is commonly believed that such quantities behave as powers of  $|p - p_c|$ , and a major open problem of percolation is to prove this. That is to say, we conjecture that the limits

$$(1.27) \quad \gamma = -\lim_{p \uparrow p_c} \frac{\log \chi(p)}{\log |p - p_c|},$$

$$(1.28) \quad \beta = \lim_{p \downarrow p_c} \frac{\log \theta(p)}{\log |p - p_c|}$$

exist, and that the limit

$$(1.29) \quad \delta^{-1} = -\lim_{n \rightarrow \infty} \frac{\log P_{p_c}(|C| \geq n)}{\log n}$$

exists, in agreement with (1.26). The quantities  $\gamma, \beta, \delta$  are called ‘critical exponents’. There are physical reasons for believing the hypothesis of ‘universality’: the numerical values of critical exponents may depend only on the dimension  $d$  and not on the structure of the particular lattice. We return to such questions in Chapters 9 and 10, where we include a summary of progress towards answers to such questions. As remarked above, substantial progress has been made under the assumption that  $d$  is sufficiently large, currently that  $d \geq 19$ .

We close this section with a review of some of the principal characters in percolation. According to one method of counting, there are four such characters:

(a) the *percolation probability*

$$\theta(p) = P_p(|C| = \infty);$$

(b) the *mean size of the open cluster at the origin*

$$\chi(p) = E_p|C|;$$

(c) the *mean size of the finite open cluster at the origin*

$$\chi^f(p) = E_p(|C|; |C| < \infty).$$

(d) The fourth such principal character is the *number of open clusters per vertex*, defined by

$$(1.30) \quad \kappa(p) = E_p(|C|^{-1}),$$

with the convention that  $1/\infty = 0$ . That is to say,

$$\kappa(p) = \sum_{n=1}^{\infty} \frac{1}{n} P_p(|C| = n).$$

We study the function  $\kappa$  in more detail in Chapter 4.

We note that

$$(1.31) \quad \chi^f(p) = \chi(p) \quad \text{whenever } \theta(p) = 0.$$

There are many useful analogies between the percolation model and the Ising model, and we note that  $\theta$  corresponds to magnetization,  $\chi^f$  to susceptibility, and  $\kappa$  to free energy per vertex.

The quantities  $\chi$ ,  $\chi^f$ , and  $\kappa$  are moments of the number of *vertices* in  $C$ . There are good reasons to define these quantities instead in terms of the number of *edges* of  $C$ , principally since such a definition would enable a unified approach to both bond and site percolation. For bond percolation on  $\mathbb{L}^d$  it matters little which route we adopt, and we have chosen that which leads to fewest technical complications later.

## 1.6 Site Percolation

There are ways of impeding flow through a medium other than blocking the *edges*, and a natural alternative is to block the *vertices* instead. The corresponding model is termed ‘site percolation’, and it is defined as follows. We designate each *vertex* of the lattice  $\mathbb{L}^d$  *open* with probability  $p$ , and *closed* otherwise; different vertices receive independent designations. A path is called *open* if all its vertices are open. The open cluster  $C(x)$  at the vertex  $x$  is defined as the set of all vertices which may be attained by following open paths from  $x$  (if  $x$  is closed, then  $C(x)$  is empty). As before, we write  $C = C(0)$ , and we define the *percolation probability*  $\theta(p) = P_p(|C| = \infty)$ , together with the *critical probability*

$$p_c = \sup\{p : \theta(p) = 0\}.$$

When we wish to emphasize the type of percolation model under study, we shall write  $\theta^{\text{site}}$  or  $\theta^{\text{bond}}$  (and  $p_c^{\text{site}}$  or  $p_c^{\text{bond}}$ ) as appropriate.

Figure 1.9 contains four snapshots of site percolation on the square lattice for different values of  $p$ . The critical probability of this process is unknown, but is believed to be around 0.59; see Section 3.1.

Most arguments available for percolation models may be adapted to both bond and site models, and for that reason we pay only little attention to site percolation in this book. Indeed, there is a sense in which every bond model may be reformulated as a site model (on a different graph); the converse is false, and therefore site models are more general than bond models. We amplify this remark next. The *covering graph* (or *line graph*) of a graph  $G$  is the graph  $G_c$  defined as follows. To each edge of  $G$  there corresponds a distinct vertex of  $G_c$ , and two such vertices are deemed adjacent if and only if the corresponding edges of  $G$  share an endvertex. Suppose we are provided with a bond percolation process on  $G$ . We call a vertex of  $G_c$  *open* if and only if the corresponding edge of  $G$  is open. This induces a site percolation process on  $G_c$ . Furthermore, it is clear that every path of open edges in  $G$  corresponds to a path of open vertices in  $G_c$  (and vice versa). [We may note that there exist site models which cannot be obtained from any bond model in the above way.]

Let us now consider an arbitrary infinite connected graph  $G = (V, E)$ . Let 0 denote a specified vertex of  $G$  which we call the ‘origin’. We define  $\theta^{\text{bond}}(p)$  (respectively  $\theta^{\text{site}}(p)$ ) to be the probability that 0 lies in an infinite open cluster of  $G$  in a bond percolation (respectively site percolation) process on  $G$  having parameter  $p$ . Clearly  $\theta^{\text{bond}}(p)$  and  $\theta^{\text{site}}(p)$  are non-decreasing functions of  $p$ , and the bond and site critical probabilities are given by

$$\begin{aligned} p_c^{\text{bond}} &= p_c^{\text{bond}}(G) = \sup\{p : \theta^{\text{bond}}(p) = 0\}, \\ p_c^{\text{site}} &= p_c^{\text{site}}(G) = \sup\{p : \theta^{\text{site}}(p) = 0\}. \end{aligned}$$

We have from the above considerations that

$$(1.32) \quad p_c^{\text{bond}}(G) = p_c^{\text{site}}(G_c).$$

It is natural to ask whether there exists a relationship between the two critical points of a given graph  $G$ .

**(1.33) Theorem.** *Let  $G = (V, E)$  be an infinite connected graph with origin 0 and maximum vertex degree  $\Delta$  ( $< \infty$ ). The critical probabilities of  $G$  satisfy*

$$(1.34) \quad \frac{1}{\Delta - 1} \leq p_c^{\text{bond}} \leq p_c^{\text{site}} \leq 1 - (1 - p_c^{\text{bond}})^\Delta.$$

One consequence of this theorem is that  $p_c^{\text{bond}}(G) < 1$  if and only if  $p_c^{\text{site}}(G) < 1$ . The third inequality of (1.34) may be improved by replacing the exponent  $\Delta$  by  $\Delta - 1$ , but we do not prove this here. Also, the strict inequality  $p_c^{\text{bond}}(G) < p_c^{\text{site}}(G)$  is valid for a broad family of graphs  $G$ ; see Section 3.4.

**Proof.** The first inequality of (1.34) follows by counting paths, as in (1.15)–(1.16). Therefore we turn immediately to the remaining two inequalities. In order to obtain these, we shall prove a certain stochastic inequality. Given two random subsets  $X, Y$  of  $V$  with associated expectation operator  $E$ , we write  $X \leq_{\text{st}} Y$ , and say that  $X$  is *stochastically dominated* by  $Y$ , if

$$E(f(X)) \leq E(f(Y))$$

for all bounded, measurable functions  $f$  satisfying  $f(A) \leq f(B)$  if  $A \subseteq B \subseteq V$ . A more systematic discussion of stochastic domination is provided in Section 7.4.

Let  $C^{\text{bond}}(p)$  be a random subset of  $V$  having the law of the cluster of bond percolation at the origin; let  $C^{\text{site}}(p)$  be a random subset having the law of the cluster of site percolation at the origin *conditional on 0 being an open vertex*. We claim that

$$(1.35) \quad C^{\text{site}}(p) \leq_{\text{st}} C^{\text{bond}}(p)$$

and that

$$(1.36) \quad C^{\text{bond}}(p) \leq_{\text{st}} C^{\text{site}}(p') \quad \text{where } p' = 1 - (1 - p)^\Delta.$$

Since

$$\begin{aligned} \theta^{\text{bond}}(p) &= P_p(|C^{\text{bond}}(p)| = \infty), \\ p^{-1}\theta^{\text{site}}(p) &= P_p(|C^{\text{site}}(p)| = \infty), \end{aligned}$$

the remaining claims of (1.34) will follow from (1.35)–(1.36). Indeed, (1.35)–(1.36) imply that

$$(1.37) \quad \frac{\theta^{\text{site}}(p)}{p} \leq \theta^{\text{bond}}(p) \leq \frac{\theta^{\text{site}}(p')}{p'} \quad \text{where } p' = 1 - (1 - p)^\Delta,$$

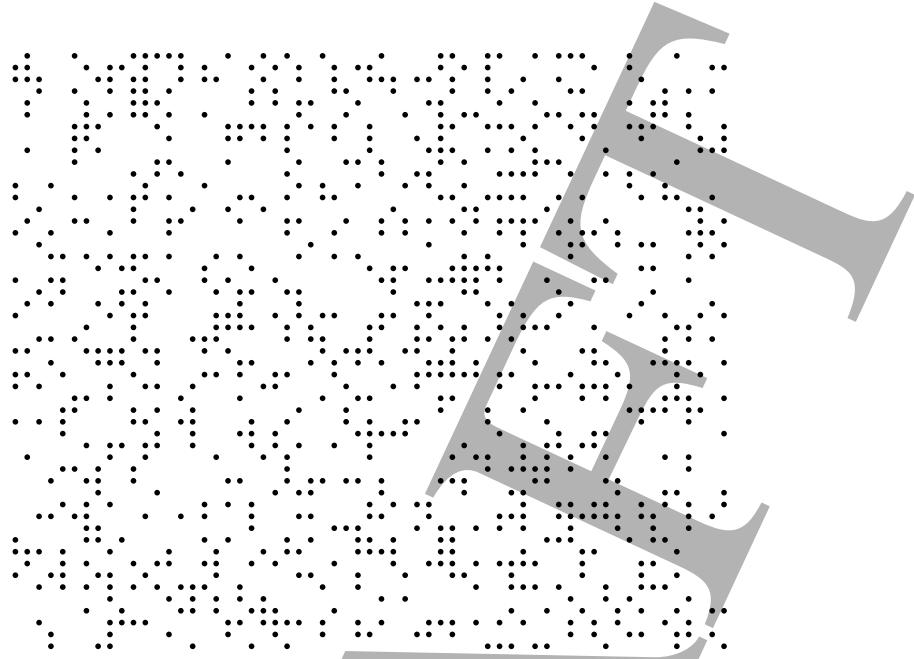
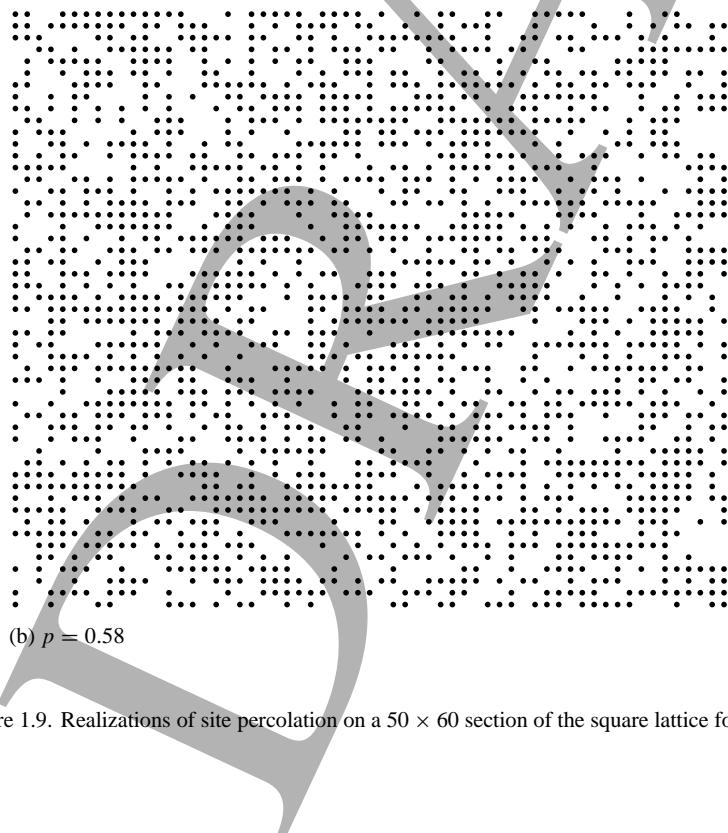
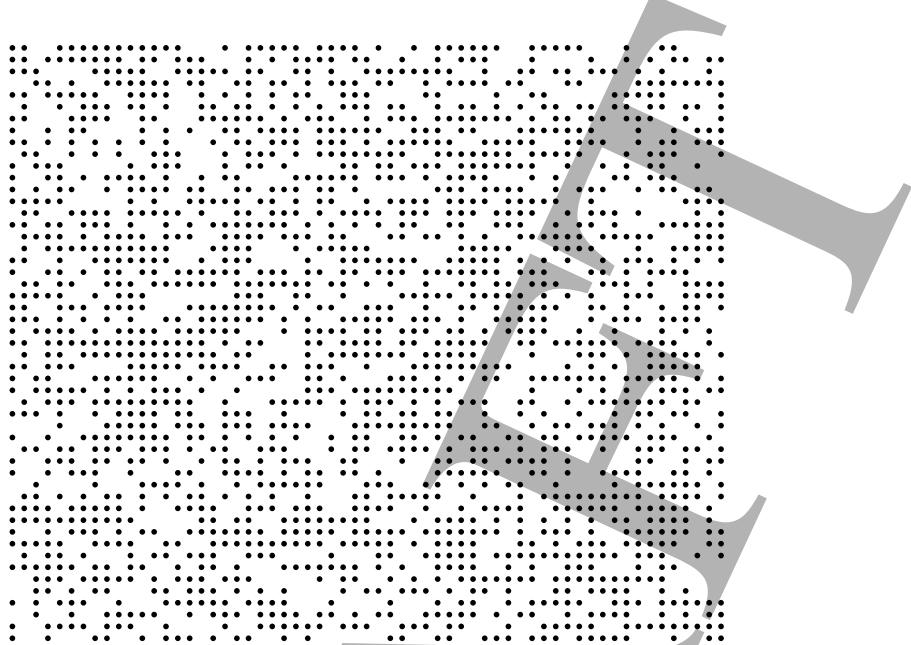
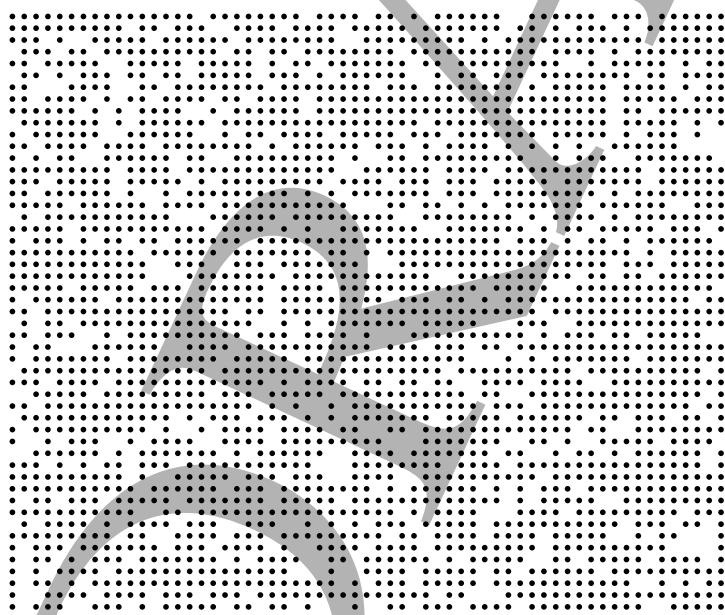
(a)  $p = 0.3$ (b)  $p = 0.58$ 

Figure 1.9. Realizations of site percolation on a  $50 \times 60$  section of the square lattice for four

(c)  $p = 0.60$ (d)  $p = 0.80$ 

different values of  $p$ . The critical value of this process is believed to be near 0.59.

which is slightly stronger than the remaining parts of (1.34).

We construct appropriate couplings of the bond and site models in order to prove (1.35)–(1.36); this is a common technique when studying two processes simultaneously, and will be used later in this book. Let  $\omega \in \{0, 1\}^E$  be a realization of a bond percolation process on  $G = (V, E)$  having density  $p$ . We may build the cluster at the origin in the following standard manner. Let  $e_1, e_2, \dots$  be a fixed ordering of  $E$ . At each stage  $k$  of the inductive construction, we shall have a pair  $(A_k, B_k)$  where  $A_k \subseteq V$ ,  $B_k \subseteq E$ . Initially we set  $A_0 = \{0\}$ ,  $B_0 = \emptyset$ . Having found  $(A_k, B_k)$  for some  $k$ , we define  $(A_{k+1}, B_{k+1})$  as follows. We find the earliest edge  $e$  in the ordering of  $E$  having the following properties:  $e \notin B_k$ , and  $e$  is incident with exactly one vertex of  $A_k$ , say the vertex  $x$ . We now set

$$(1.38) \quad A_{k+1} = \begin{cases} A_k & \text{if } e \text{ is closed,} \\ A_k \cup \{y\} & \text{if } e \text{ is open,} \end{cases}$$

$$(1.39) \quad B_{k+1} = \begin{cases} B_k \cup \{e\} & \text{if } e \text{ is closed,} \\ B_k & \text{if } e \text{ is open,} \end{cases}$$

where  $e = \langle x, y \rangle$ . If no such edge  $e$  exists, we declare  $(A_{k+1}, B_{k+1}) = (A_k, B_k)$ . The sets  $A_k, B_k$  are non-decreasing, and the open cluster at the origin is given by  $A_\infty = \lim_{k \rightarrow \infty} A_k$ .

We now augment the above construction in the following way. We colour the vertex 0 *red*. Furthermore, on obtaining the edge  $e$  given above, we colour the vertex  $y$  *red* if  $e$  is open, and *black* otherwise. We specify that each vertex is coloured at most once in the construction, in the sense that any vertex  $y$  which is obtained at two or more stages is coloured in perpetuity according to the first colour it receives.

Let  $A_\infty(\text{red})$  be the set of points connected to the origin by red paths of  $G$  (that is, by paths all of whose vertices are red). We make two claims concerning  $A_\infty(\text{red})$ :

- (i) it is the case that  $A_\infty(\text{red}) \subseteq A_\infty$ , and all neighbours of vertices in  $A_\infty(\text{red})$  which do not lie in  $A_\infty(\text{red})$  are black;
- (ii)  $A_\infty(\text{red})$  has the same distribution as  $C^{\text{site}}(p)$ ;

and inequality (1.35) follows immediately from these claims.

Claim (i) is straightforward. In order to be coloured red, a vertex is necessarily connected to the origin by a path of open edges. Furthermore, since all edges with exactly one endvertex in  $A_\infty$  are closed, all neighbours of  $A_\infty(\text{red})$  are necessarily black.

We sketch an explanation of claim (ii). Whenever a vertex is coloured either red or black, it is coloured red with probability  $p$ , independently of all earlier colourings. This is not a full proof of (ii) but will satisfy many readers. More details are provided by Grimmett and Stacey (1998); see also the proof of Lemma (3.29).

The derivation of (1.36) is similar. We start with a directed version of  $G$ , namely the directed graph  $\vec{G} = (\vec{V}, \vec{E})$  obtained from  $G$  by replacing each edge

$e = \langle x, y \rangle$  by two directed edges, one in each direction, and denoted respectively by  $[x, y]$  and  $[y, x]$ . We now let  $\vec{\omega} \in \{0, 1\}^{\vec{E}}$  be a realization of an (oriented) bond percolation process on  $\vec{G}$  having density  $p$ .

We colour the origin *green*. We colour a vertex  $x (\neq 0)$  *green* if at least one edge  $f$  of the form  $[y, x]$  satisfies  $\vec{\omega}(f) = 1$ ; otherwise we colour  $x$  *black*. Then

$$(1.40) \quad P_p(x \text{ is green}) = 1 - (1 - p)^{\rho(x)} \leq 1 - (1 - p)^\Delta,$$

where  $\rho(x)$  is the degree of  $x$ , and  $\Delta = \max_x \rho(x)$ .

We now build a copy  $A_\infty$  of  $C^{\text{bond}}(p)$  more or less as described in (1.38)–(1.39). The only difference is that, on considering the edge  $e = \langle x, y \rangle$  where  $x \in A_k$ ,  $y \notin A_k$ , we declare  $e$  to be open for the purpose of (1.38)–(1.39) if and only if  $\vec{\omega}([x, y]) = 1$ . Finally, we let  $A_\infty(\text{green})$  be the set of points connected to the origin by green paths. It may be seen that  $A_\infty(\text{green}) \supseteq A_\infty$ . Furthermore, by (1.40),  $A_\infty(\text{green})$  is stochastically dominated by  $C^{\text{site}}(p')$  where  $p' = 1 - (1 - p)^\Delta$ . Inequality (1.36) follows.  $\square$

## 1.7 Notes

**Section 1.1.** The origins of the mathematical theory of percolation may be found in the work of Flory (1941), Broadbent (1954), and Broadbent and Hammersley (1957). Hammersley (1983) and Grimmett (1999a) have described something of the history of the subject. We know of four books to date, the serious mathematical text of Kesten (1982), the gentle account by Efros (1986), and the books of Stauffer and Aharony (1991) and Hughes (1996). Of the many reviews, we mention Chayes and Chayes (1986a), Menshikov, Molchanov, and Sidorenko (1986), Aizenman (1987), Grimmett (1987b, 1997), Kesten (1987e), and Newman (1987a).

For general discussions of periodic lattices in  $d$  dimensions, see Grimmett (1978a, b), Godsil and McKay (1980), and Kesten (1982, Chapters 2, 3). Following Grimmett and Newman (1990), there has been serious study of percolation on graphs whose growth functions grow faster than any polynomial. Many such graphs arise as Cayley graphs of groups, and a systematic study of such systems has been initiated. See Benjamini and Schramm (1996), Benjamini, Lyons, Peres, and Schramm (1997), and the references therein.

The reader is referred to Chapter 12 for references appropriate to mixed, inhomogeneous, long-range, oriented, and first-passage percolation. The relationship between percolation and other models of statistical physics has been explored by Essam (1972, 1980); see also Section 13.6. Wierman (1987b) has studied ‘high density’ percolation in two dimensions; in this percolation-type process, one is interested only in those vertices which are incident to at least  $k$  open edges, for some specified value of  $k$ .

**Section 1.3.** The device for defining all percolation processes on the same probability space seems to have appeared for the first time in Hammersley (1963).

**Section 1.4.** The existence of a critical phenomenon was observed by Broadbent and Hammersley (1957) and Hammersley (1957a, 1959). Some people say that ‘percolation occurs’ when there exists an infinite open cluster.

Bond percolation on the line  $\mathbb{Z}$  is essentially the problem of ‘runs’; see Feller (1968). The proof that  $p_c(2) = \frac{1}{2}$  was performed by Kesten (1980a), who built upon earlier arguments of Harris (1960), Russo (1978), and Seymour and Welsh (1978). Wierman (1981) adapted Kesten’s proof to calculate exact values for critical probabilities for bond percolation on the triangular and hexagonal lattices.

Kesten (1982) proved that  $p_c(3) < p_c(2)$ ; see also the related work of Menshikov, Molchanov, and Sidorenko (1986, Section 4). J. van den Berg (unpublished) and A. Frieze (unpublished) have pointed out a simple argument which leads to the strict inequality  $p_c(d+1) < p_c(d)$  for  $d \geq 1$ . The argument amounts to the following for  $d = 2$ . Each edge of  $\mathbb{L}^2$  may be thought of as the bottom of two infinite ladders of  $\mathbb{L}^3$ . By subdividing the ‘vertical’ edges in such ladders, we may construct disjoint ladders above different edges. By choosing the edge-probabilities for the subdivided edges with care, we arrive at the conclusion that  $p_c(3) \leq 0.4798$ , whereas  $p_c(2) = \frac{1}{2}$ .

Menshikov (1987a) has provided a more general argument (see also Zuev (1987)), showing that  $p_c(\mathcal{L}_1) < p_c(\mathcal{L}_2)$  whenever  $\mathcal{L}_2$  is a strict sublattice of  $\mathcal{L}_1$  satisfying certain conditions; he is able to find non-trivial lower bounds for the difference  $p_c(\mathcal{L}_2) - p_c(\mathcal{L}_1)$ . His argument may be adapted to obtain a canonical approach to proving strict inequalities for general processes and enhancements thereof. See Aizenman and Grimmett (1991) and Sections 3.2–3.3 of the present book.

Kesten (1990) has proved that  $p_c(d) \sim (2d)^{-1}$  as  $d \rightarrow \infty$ ; see also Gordon (1991), Kesten (1991), and Kesten and Schonmann (1990). The detailed asymptotics of the forthcoming equation (3.2) were presented by Hara and Slade (1995). See also Hughes (1996).

The duality of planar lattices was explored by Hammersley (1959) and Harris (1960), and later by Fisher (1961). For an account of self-avoiding walks and the connective constant, see Hammersley (1957b); a recent account has been provided by Madras and Slade (1993).

**Section 1.5.** We defer the list of references until the appropriate forthcoming chapters.

**Section 1.6.** It was remarked by Fisher (1961) and Fisher and Essam (1961) that a bond model may be transformed into a site model; see also Kesten (1982, Chapter 3). We shall see at Theorem (2.8) that the definitions of  $p_c^{\text{bond}}(G)$  and  $p_c^{\text{site}}(G)$  are independent of the choice of origin, whenever the graph  $G$  is connected. Theorem (1.33) appeared in Grimmett (1997). The exponent  $\Delta$  in the final inequality of (1.34) has been improved to  $\Delta - 1$  by Grimmett and Stacey (1998), where

one may find also the strict inequality  $p_c^{\text{bond}}(G) < p_c^{\text{site}}(G)$  for a broad class of graphs  $G$  including all finite-dimensional lattices in two or more dimensions. Further information concerning strict inequalities may be found in Chapter 3, and particularly in Section 3.4.

