The Annals of Probability
1987, Vol. 15, No. 4, 1231–1271

THE 1986 WALD MEMORIAL LECTURES

PERCOLATION THEORY AND FIRST-PASSAGE PERCOLATION[1]

BY HARRY KESTEN

Cornell University

We give a survey of some of the principal results in percolation theory and first-passage percolation. No proofs are given. Open problems abound and are usually displayed with a number (∗P).

Table of contents

0. General notation and conventions. This section is for reference only. We include some standard notation not defined in the text.

C_i will stand for a strictly positive, finite constant, whose precise value is of little importance for us. In fact, C_i may have different values at different appearances.

We call a function f *increasing* if $x_1 \geq x_2$ implies $f(x_1) \geq f(x_2)$, and *strictly increasing* if $x_1 > x_2$ implies $f(x_1) > f(x_2)$.

For $v = (v(1), \dots, v(d)) \in \mathbb{R}^d$,

$$|v| = \max\{|v(i)|: 1 \leq i \leq d\},$$

$$\|v\|_2 = \left\{\sum |v(i)|^2\right\}^{1/2}.$$

Received September 1986.
[1]Research supported by the NSF through a grant to Cornell University.
AMS 1980 *subject classifications.* Primary 60K35, 82A43; secondary 82A68, 60K10.
Key words and phrases. Percolation, first-passage percolation, phase transition, power laws, critical exponents, scaling relations, resistance of random networks, flow problem, incipient infinite cluster, invasion percolation, asymptotic shape, Eden's model, time constants, minimal surfaces.

For two strictly positive functions f and g

$$f \approx g \text{ means } \frac{\log f}{\log g} \to 1,$$

as a suitable argument approaches a special value.

The indicator function of a set or event A is denoted by $I[A]$.

∂ denotes the topological boundary. Thus, $\partial([-n, n]^d) = \{x \in \mathbb{R}^d: |x(j)| \le n, 1 \le j \le d$, but $|x(i)| = n$ for some $1 \le i \le d\}$. P_p is the product probability measure on the configuration space with bonds (or sites) open with probability p. P_{cr} stands for P_{p_c}.

1. Description of the simplest percolation model and first problems. Percolation theory as known today originated with Broadbent and Hammersley [11]. A precursor of the model appears in Flory [39]. Broadbent and Hammersley wanted to model the spread of a fluid (or gas) through a random medium. The difficulty was to describe the random medium; the randomness in their model is totally in the medium and not (as in the usual diffusion model) in the motion of the fluid. Broadbent [10] already modeled the medium as a system of channels, some of which are wide and others narrow. He assumed that the fluid passes through all the wide channels, but not through any of the narrow ones. The system of channels was represented in idealized form by the edges of $\mathbb{Z}^d$. We take $d \ge 2$ throughout, the case $d = 1$ being trivial. One now takes each of the edges, independently of all others, *open* (or *passable*) with probability p and *closed* (or *blocked*) with probability $q := 1 - p$. The corresponding probability measure on the configurations of open and closed edges is denoted by P_p; E_p denotes expectation with respect to P_p. If the origin $\mathbf{0}$ is a source of fluid, which points are wetted? Denote the set of these points by W. Thus,

(1.1) $W =$ collection of points connected to $\mathbf{0}$ by an open path on $\mathbb{Z}^d$.

Here, and in the sequel an *open path* is a path all of whose edges are open. We use $A \to B$ to denote that there exists an open path from some vertex in A to some vertex in B. W is called the *open cluster* of $\mathbf{0}$, and the open cluster of a vertex v is defined by replacing $\mathbf{0}$ by v in (1.1). Most questions in percolation theory concern some aspect of the distribution of W. Despite its simplicity the model is very rich, as I hope to demonstrate. I find it fascinating that this very elementary and easily explained model leads to a large variety of quite difficult (and many still unsolved) problems. The first question raised by Broadbent and Hammersley is: What is the probability that points very far out are wetted? In the limit the question becomes, "What is

$$\theta(p) := P_p\{W \text{ is infinite}\}?\text{"}$$

$\theta(p)$ is called the *percolation probability*. We do not know how to calculate $\theta(p)$, but it is easy to see that

(1.2) $\theta(0) = 0, \quad \theta(1) = 1 \quad \text{and} \quad p \to \theta(p)$ is increasing.

(The last property follows, at least intuitively, from the fact that increasing p

makes it more likely to have open edges, and hence more likely to have an open path from **0** to infinity.) One can therefore define the *critical probability*

$$p_c := \sup\{p: \theta(p) = 0\}. \tag{1.3}$$

[In several of the references this critical probability is denoted by p_H to distinguish it from p_T of (1.9). Since it is now known (cf. the answer to (1.10)) that $p_T = p_H$, the common notation p_c seems preferable.] By definition

$$\theta(p) = P_p\{W \text{ is infinite}\} \begin{cases} = 0, & \text{for } p < p_c, \\ > 0, & \text{for } p > p_c. \end{cases} \tag{1.4}$$

The general shape of the graph of $\theta(\cdot)$ is believed to be as in Figure 1. The

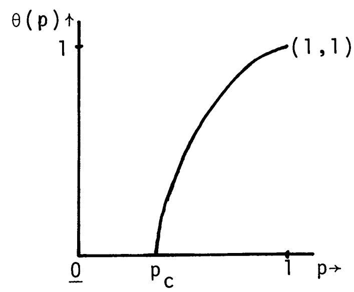

FIG. 1. *General features of the graph of* $\theta(\cdot)$.

subject came to life with Broadbent and Hammersley's proof in [11] and [50] that $0 < p_c < 1$, so that there are two regimes with totally different global behavior of the system. For $p < p_c$ there are no infinite clusters, while simple ergodic considerations (see [54]) show that for $p > p_c$ there are w.p. 1 infinite clusters (or "percolation occurs"). In fact, a recent result of Aizenman, Kesten and Newman [4] shows that there is w.p. 1 a *unique* infinite cluster, whenever $\theta(p) > 0$. The system is said to undergo *phase transition* at p_c. This is one of the main reasons of the popularity of the subject these days. The popularity can be measured by the fact that the surveys and books [13], [25], [27], [33], [34], [36], [53], [59], [64], [74], [89], [90], [91] and [97], all of which appeared in or after 1979, devote about 1700 pages to percolation. Statistical physicists hope that the model is simple enough to demonstrate explicitly many aspects of phase transitions or critical phenomena known experimentally or (often on a nonrigorous basis) from other models. There is even a rigorous relation between the Ising model—or more generally the Potts model—for magnetism and percolation due to Fortuin and Kasteleyn [57] and [40] (see also [35], Section 6A). The heuristic interplay between these models has, however, been much greater than one would expect on the basis of this rigorous relation alone; results in percolation have often stimulated a similar result in the Ising model and vice versa. We list some of the questions suggested by considerations of statistical mechanics or immediate mathematical curiosity and indicate the answers as far as known today.

(1.5P) What is p_c?

For $d = 2$, $p_c = \frac{1}{2}$ [59], but p_c is still unknown for $d \geq 3$.

(1.6P) Is $p \to \theta(p)$ continuous?

Yes, at all $p \neq p_c$ ([4]), and for $d = 2$ also at p_c ([54], [85] and [59], Theorem 3.1). One believes, but has not yet proved, that $\theta(\cdot)$ is continuous at p_c also for $d \geq 3$. Note that by (1.4) continuity at p_c is equivalent to $\theta(p_c) = 0$, or to the absence of infinite open clusters at p_c. This is not the case in all generalizations of the model (cf. [6]).

Several new questions arise when one tries to describe more details of the distribution of

$$\#W := \text{number of edges in } W.$$

Note that $\theta(p)$ is simply the atom at ∞ of this distribution. For instance, one can consider

$$\chi(p) := E_p\{\#W;\ \#W < \infty\} \tag{1.7}$$

and

$$E_p\{\#W\} = \chi(p) + \theta(p) \cdot \infty, \tag{1.8}$$

and define

$$p_T := \sup\{p\colon E_p\{\#W\} < \infty\}. \tag{1.9}$$

Clearly, $p_T \leq p_c$, since by definition $E_p\{\#W\} \geq \theta(p) \cdot \infty = \infty$ for $p > p_c$. Some authors did not distinguish between p_T and p_c, but it became clear from [84] and [86] that this is an important distinction. In fact, the latter paper reduced the proof of $p_c = \frac{1}{2}$ for $d = 2$ to proving $p_T = p_c$ for $d = 2$. One thus has the general question:

(1.10) Is $p_T = p_c$ for all d?

Menshikov, Molchanov and Sidorenko [74], and independently Aizenman and Barsky [1], have just shown that the answer to (1.10) is *yes*. This is even true in much more general models. The results of [1] and [74] immediately imply $p_c = \frac{1}{2}$ for the above model in dimension 2 and also provide the main step for rederiving the known critical probabilities in other models, to be discussed in the next section. Moreover, in contrast to the usual approach, their method uses little geometry. We therefore consider [74] and [1] as the most important and exciting recent papers on the subject. Other obvious questions are:

(1.11P) How do $P_p\{\#W = n\}$ and $P_p\{\text{Rad}(W) = n\}$ or $P_p\{n \leq \#W < \infty\}$ and $P_p\{n \leq \text{Rad}(W) < \infty\}$ behave, as functions of n and/or p?

where

$$\text{Rad}(W)\ (= \text{radius of } W) := \max\{|v|\colon v \in W\}. \tag{1.12}$$

Not surprisingly, the asymptotic behavior in n of the above functions is quite different for $p < p_c$, $p = p_c$ and $p > p_c$, and only partial information is available.

Consider first $p < p_c$. Subadditivity arguments as in [3] and [59], Lemma 5.9, show that (for any p)

$$\lim \frac{1}{n} \log P_p\{\#W = n\} \text{ exists.} \tag{1.13}$$

However, the limit in (1.13) is nontrivial, i.e., finite and strictly positive, only if $p < p_c$. Some partial information about this limit can be obtained from [5], [59], Theorem 5.1, and the forthcoming article by Nguyen [77].

Grimmett (private communication) pointed out that simple subadditivity arguments show that for $p < p_c$, there exists a $\xi(p) = \xi(p, d) \in (0, \infty)$, such that

$$\begin{aligned} C_1 n^{1-d} \exp\left(-\frac{n}{\xi(p)}\right) &\le P_p\{\mathrm{Rad}(W) \ge n\} \\ &\le C_2 n^{d-1} \exp\left(-\frac{n}{\xi(p)}\right). \end{aligned} \tag{1.14}$$

The quantity $\xi(p)$ is the same as the *correlation length*, as defined in Chayes, Chayes and Fröhlich [18] (see also [13], Section 2.3) for $p < p_c$. More specifically, if we define

$$\tau(p, x, y) := P_p\{x \to y;\ \#W < \infty\}, \tag{1.15}$$

then for $p < p_c$ (where $\#W < \infty$ is automatic)

$$\frac{1}{\xi(p)} = \lim_{n \to \infty} -\frac{1}{n} \log \tau(p, \mathbf{0}, (n, 0, \dots, 0)), \tag{1.16}$$

and (since τ is supermultiplicative)

$$\tau(p, \mathbf{0}, (n, 0, \dots, 0)) \le \exp\left(-\frac{n}{\xi(p)}\right), \qquad p < p_c. \tag{1.17}$$

The function τ in (1.15) is sometimes called the *truncated connectivity function*. The ordinary *connectivity function* is $P_p\{x \to y\}$.

The exponential factor in (1.14) and (1.17) does not becomes significant until n is of the order of the correlation length. For n much larger than the correlation length, the probability of an open path between points a distance n apart becomes small when $p < p_c$. On the other hand, it is known (by [59], Theorem 5.1, and the result $p_T = p_H$; see also [85]) that at p_c the probability of an open crossing of $[0, n] \times [0, 3n]^{d-1}$ in the first coordinate direction is bounded away from 0 (as $n \to \infty$). Thus, for $p < p_c$ one can think of the correlation length as the smallest scale at which the connectivity probabilities differ drastically from those at p_c. This intuitive interpretation of correlation length is also useful for $p > p_c$ and is also a heuristic support for scaling theory (see the end of Section 2.3). So far we only have a way of making this precise if $d = 2$ (cf. [66]). For fixed

n, $\tau(p, \mathbf{0}, (n, 0, \ldots, 0))$ and $P_p\{\text{Rad}(W) \geq n\}$ are continuous in p, so that the above intuitive interpretation also tells us that we must have

$$(1.18) \qquad \xi(p) \uparrow \infty, \quad \text{as } p \uparrow p_c.$$

In fact, this follows from the fact that $\xi(p)$ in (1.14) and (1.16) is increasing on $[0, p_c)$, but that $\tau(p_c, \mathbf{0}, (n, 0, \ldots, 0))$ and $P_{\text{cr}}\{\text{Rad}(W) \geq n\}$ are known not to decrease exponentially in n [cf. (1.21) and (1.22)]. Here, and below, P_{cr} stands for P_{p_c}.

For $p = p_c$ it is believed that

$$(1.19\text{P}) \qquad P_{\text{cr}}\{\#W \geq n\} \approx n^{-1/\delta},$$

for some $0 < \delta < \infty$, in the sense that

$$(1.20) \qquad (\log n)^{-1} \log P_{\text{cr}}\{\#W \geq n\} \to -1/\delta.$$

However, it is only known for general d that

$$(1.21) \qquad \begin{aligned} P_{\text{cr}}\{\#W \geq n\} &\geq P_{\text{cr}}\{\text{Rad}(W) \geq n\} \\ &\geq C_1 n^{-(d-1)/2} \end{aligned}$$

(e.g., by the method of [94], Corollary 3.15, combined with [59], Theorem 5.1). Also (cf. [1])

$$(1.22) \qquad \sum_{n=1}^{\infty} P_{\text{cr}}\{\#W \geq n\} e^{-nh} \geq C_2 h^{-1/2},$$

which is equivalent to

$$P_{\text{cr}}\{\#W \geq n\} \geq C_3 n^{-1/2}.$$

For $d = 2$ we have the more detailed estimates (cf. [59], Chapter 8, Kesten in [64], [66], [84], [86] and [87], Chapter 3)

$$(1.23) \qquad \begin{aligned} C_4 n^{-1/3} \leq P_{\text{cr}}\{\text{Rad}(W) \geq n\} &\leq P_{\text{cr}}\{\#W \geq n\} \\ &\leq C_5 n^{-C_6}. \end{aligned}$$

For $p = p_c$ one also expects

$$(1.24\text{P}) \qquad P_{\text{cr}}\{\text{Rad}(W) \geq n\} \approx n^{-1/\delta_r},$$

for some $0 < \delta_r < \infty$, but so far one can only prove (1.21)–(1.23).

As discussed in [5] it would be very desirable to have an answer to the following problem when $d \geq 3$:

(1.25P) Find a good upper bound for the left-hand side of (1.21) and for $\tau(p_c, \mathbf{0}, x)$.

Finally, for $p > p_c$ it is believed that

$$(1.26\text{P}) \qquad \log P_p\{\#W = n\} \sim -C_7 n^{(d-1)/d}.$$

Lower bounds of this form for the left-hand side are in [3] and [59], Theorem 5.2, and upper bounds in [20].

2. Generalizations and applications.

2.1. *Bond and site percolation on general graphs.* To motivate generalizations and further problems we begin with the following application or interpretation of percolation, due to Frisch and Hammersley [41]. Think of the vertices of $\mathbb{Z}^2$ as the trees in an orchard and assume that a blight starts at the tree at $\mathbf{0}$. We now want to use percolation as a model for the spread of the blight through the orchard. The model of Section 1, which is called (Bernoulli) *bond percolation* on $\mathbb{Z}^d$, might be reasonable if the event of a tree at v infecting a tree at a neighbor w depends only on the soil between v and w. If, on the other hand, this event depends only on the resistance to the blight of the tree at w, then a better model would be one in which the randomness is attached to the sites or vertices, rather than the bonds. This leads to *site percolation*, in which each site is independently "occupied" with probability p and "vacant" with probability q. Clearly, one can also formulate mixed site–bond problems, and one may replace $\mathbb{Z}^d$ by any graph $\mathscr{G}$ periodically imbedded in $\mathbb{R}^d$ (cf. [59], Section 2.1). E.g., a favorite graph $\mathscr{G} \subset \mathbb{R}^2$ is the triangular lattice of Figure 2. The reader will have no

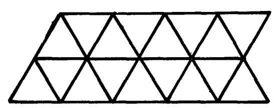

FIG. 2. *The triangular lattice.*

trouble finding his own interpretation for the two states of the bonds or sites. When considering alloys of materials A and B, "occupied" or "open" may mean "occupied by component A" and "vacant" or "closed" may mean "occupied by component B." In the case of electron spins the states may be "spin up" and "spin down," while there are still other interpretations in the resistance problems of Section 2.5.

It can be shown that bond percolation on $\mathscr{G}$ is equivalent to site percolation on the covering graph or line graph of $\mathscr{G}$ (cf. [59], Sections 2.5 and 3.1), but not vice versa, so that site percolation is somewhat more general than bond percolation, as long as one restricts oneself to models in which the sites (or bonds) are independent of each other. Percolation with the latter property is usually called *Bernoulli percolation.*

For various models the independence assumption is inappropriate. Consider, for instance, the following simple-minded model of a forest fire or spread of an epidemic (see, for instance, [70]). Again the vertices of $\mathbb{Z}^2$ are the trees. If a tree at v is ignited it burns for a random time $\lambda = \lambda(v)$ during which it may ignite some of its neighbors which are not yet burning or have not yet burned. For fixed v the events $\{v \text{ ignites } w\colon w \text{ a neighbor of } v\}$ are usually dependent since the probability that v ignites w will be an increasing function of $\lambda(v)$. Thus, even for i.i.d. $\lambda(v)$'s, the occurrence of the event v ignites w_1 will tell us something about $\lambda(v)$, which in turn influences whether v ignites w_2. Dependence destroys the simplicity of the model and *we consider here only Bernoulli percolation.*

The last example also illustrates the need to consider directed graphs $\mathscr{G}$; $\{v \text{ ignites } w\}$ should be distinguished from $\{w \text{ ignites } v\}$. Thus, in general, we want to consider one edge directed from v to w and another from w to v. Of course, one or both of these edges may be absent, and they could be dependent. E.g., the original model of bond percolation on $\mathbb{Z}^d$ can be viewed as one in which each pair of adjacent vertices is connected by a pair of edges with opposite orientation, and both edges are open (closed) simultaneously with probability p (q, respectively). *Directed or oriented percolation* discusses many of the problems of ordinary percolation, but has also some new problems. We shall not discuss this generalization here. A nice survey is in Durrett [27]; see also Durrett and Schonmann in [64].

As a final generalization we mention *multiparameter problems*. In the above examples of the spread of the blight or forest fire, one may want to treat the horizontal and vertical bonds of $\mathbb{Z}^2$ differently, for instance, because of prevailing wind conditions. Thus, one may want to take horizontal and vertical bonds open with probabilities p_{hor} and p_{ver}, respectively. It is obvious how to phrase a general multiparameter model on a graph $\mathscr{G}$ (cf. [59], Section 3.2). All the functions introduced so far now become functions of several parameters and most problems have immediate multiparameter versions. E.g., (1.5) is replaced by, "For what parameter values do infinite open clusters occur?" Many more variants appear in the literature (see, for instance, Halley in [25] and Wierman in [64]), but to keep the length of this review down we restrict ourselves to one-parameter Bernoulli percolation, and with the exception of Sections 2.2–2.4, even to bond percolation on $\mathbb{Z}^d$.

2.2. *Power laws and scaling relations.* Consider one-parameter site percolation on a periodic graph $\mathscr{G}$ in $\mathbb{Z}^d$ (as pointed out above, bond percolation is a special case of this). The critical probability p_c depends on $\mathscr{G}$. Not too much effort is devoted these days to finding $p_c(\mathscr{G})$, even though $p_c(\mathscr{G})$ is unknown with few exceptions. E.g., $p_c = \frac{1}{2}$ for site percolation on the triangular lattice ([59], pages 52–54) and $p_c = 2\sin(\pi/18)$ for bond percolation on the triangular lattice ([96]); but p_c is not known for site percolation on $\mathbb{Z}^d$, $d \geq 2$, or bond percolation on $\mathbb{Z}^d$, $d \geq 3$. The reason for the limited interest in this problem is that $p_c(\mathscr{G})$ seems to depend too much on special properties of $\mathscr{G}$. Ad hoc methods which work for one $\mathscr{G}$ seem to tell us nothing about another graph. Instead, most efforts seem to be devoted to so called *power laws* and *critical exponents*, which are supposed to be *universal* in a sense to be explained now.

Many of the quantities introduced previously, have a singularity at p_c, when considered as a function of p. This singularity is supposed to have the form of a power law, that is, the function or one of its derivatives is believed to blow up like a power of $|p - p_c|$, as in (2.4)–(2.6). Define

$$\begin{aligned} \Delta(p) &= \sum_{n=1}^{\infty} \frac{1}{n} P_p\{\#W = n\} \\ &= E_p\{(\#W)^{-1};\ \#W \geq 1\}. \end{aligned} \tag{2.1}$$

$\Delta(p)$ is the average number of clusters per site and was first considered by Sykes and Essam [92] in an attempt to calculate $p_c(\mathscr{G})$ for various $\mathscr{G}$ (see also [46] and [95]). This function is an analogue of free energy in statistical mechanics (cf. [57], [40] and [35], Section 6A). This analogy has been useful, at least in a heuristic sense, to a number of people. E.g., [1] and [4] use this function heavily. Sykes and Essam's [92] arguments were based on the assumption that $\Delta(\cdot)$ has exactly one singularity and that this is located at p_c. It is still not known what this singularity is, if any, but one generally believes that

$$\Delta'''(p) \approx |p - p_c|^{-1-\alpha}, \quad \text{for some } -1 < \alpha < 0. \tag{2.2}$$

As in (1.20), (2.2) is to be interpreted as

$$\lim_{p \to p_c} (\log|p - p_c|)^{-1} \log \Delta'''(p) = -1 - \alpha, \tag{2.3}$$

even though it is possible that a much stronger asymptotic relation between the two sides of (2.2) holds. Similar comments apply to (2.4)–(2.7). Some other proposed power laws are

$$\theta(p) \approx (p - p_c)^{\beta}, \qquad p > p_c, \tag{2.4}$$

$$\chi(p) \approx |p - p_c|^{-\gamma}, \tag{2.5}$$

$$\xi(p) \approx |p - p_c|^{-\nu}. \tag{2.6}$$

[This requires defining the correlation length also for $p > p_c$, since it is not clear that (1.16) will work for $p > p_c$. A further discussion of (2.6) for $d = 2$ is in [66].]

Still other power laws refer to behavior of functions of n, with p fixed at p_c, rather than behavior of functions of p. We already mentioned (1.19) and (1.24), and an additional one is

$$\tau(p_c, v, w) \approx |v - w|^{-(d-2+\eta)}. \tag{2.7}$$

It is conceivable that one may have to distinguish the approach of p to p_c from the positive side and from the negative side, i.e., one may have to replace (2.3) by the two relations

$$\lim_{p \downarrow p_c} (\log|p - p_c|)^{-1} \log \Delta'''(p) = -1 - \alpha_+,$$

$$\lim_{p \uparrow p_c} (\log|p - p_c|)^{-1} \log \Delta'''(p) = -1 - \alpha_-,$$

with distinct α_+ and α_-, and similarly for (2.5) and (2.6). The available evidence, numerical from simulations and theoretical ([66]), points in the direction of equality of the limits for $p \downarrow p_c$ and $p \uparrow p_c$.

The $\alpha, \beta, \gamma, \nu, \eta, \delta, \delta_r$ are called *critical exponents* and (with the exception of α) are believed to be strictly positive numbers. The *universality* conjecture states that within one class of models these exponents depend only on the dimension d, but not on the remaining structure of $\mathscr{G}$ (as long as $\mathscr{G}$ has only a bounded number of vertices in each unit cell and a bounded maximal coordination number and perhaps a few symmetry properties). Thus, for instance, the exponents for bond and site percolation on $\mathbb{Z}^2$ and the triangular lattice of Figure 2 should all be the same. However, one will get different exponents for oriented percolation on these lattices.

Other important conjectures concerning the critical exponents are the *scaling relations*. The exponents introduced so far are not supposed to be all independent, but expressible in terms of two of them. One form of the conjectured scaling relations is as follows:

$$2 - \alpha = \gamma + 2\beta = \beta(\delta + 1), \tag{2.8}$$

$$2 - \eta = d\frac{\delta - 1}{\delta + 1}, \tag{2.9}$$

$$\delta = d\delta_r - 1, \tag{2.10}$$

(2.11) all members of (2.8) equal $d\nu$.

Actually, relations (2.9)–(2.11), which involve d, are usually referred to as *hyperscaling* relations and are only supposed to hold for d less than or equal to a so called *upper critical dimension*, which is believed to be 6 (see below). For $d \geq 6$ one believes that the exponents no longer depend on d, but take on their so-called *mean field value*, which is the value these exponents have when $\mathscr{G}$ is a tree of a fixed finite coordination number (a Bethe tree). (The mean field value for η has to be defined in a different way, though.) In particular, if $\mathscr{G}$ is a binary tree all the power laws hold and the exponents can be calculated. One often thinks of such a tree as corresponding to $\mathbb{Z}^d$ for infinite d, by thinking of each pair of edges between an nth generation vertex and its children as being translates of the positive and negative coordinate vectors of a separate dimension. Thus, the upper critical dimension conjecture can be paraphrased as "for any $d \geq 6$ one has the same behavior as in infinite dimensions." The value 6 for the upper critical dimension was guessed by Toulouse [93], because substitution of the mean field values for α, β, γ, δ, η and ν yield equalities in (2.8)–(2.11) when one takes $d = 6$.

Very little of this *scaling theory* has been proven so far. For many exponents one has been able to prove that the mean field value is a one-sided bound for the exponent for all $d \geq 2$ (cf. [5] for γ, [14] for β and [1] for δ). One also has bounds in dimension 2 which show that most exponents do not yet take their mean field value for bond or site percolation on $\mathbb{Z}^2$ (cf. [66] and [67]). For $d = 2$ specific rational values have even been conjectured for the exponents (cf. [90] and the references therein).

Even though none of the power laws has been made rigorous, it has been proven in [66] for bond and site percolation on $\mathbb{Z}^2$ that once we have the two power laws (1.24) and (2.6), then the other power laws and all the scaling relations (2.8)–(2.11) with $d = 2$ must hold. There is one exception. We cannot prove the first equality in (2.8); very little is known about α and it is not even known yet that $\Delta'''(p)$ has a singularity at p_c. It should be noted that the results of [66] do not rely on (1.24) and (2.6) as assumptions. Independently of these, [66] derives relations between various quantities which are interesting by themselves. For example, for bond and site percolation on $\mathbb{Z}^2$ it is shown that

$$\frac{\chi(p)}{[\xi(p)P_{\mathrm{cr}}\{\mathrm{Rad}\, W \geq \xi(p)\}]^2}$$

is bounded away from 0 and ∞ as $p \to p_c$. When this relation is combined with (1.24) and (2.6), one obtains (2.5) plus one of the scaling relations.

Durrett and Nguyen [28], [30] and [76] have proven a number of inequalities between critical exponents in all dimensions, provided the critical exponents exist. Typically, these inequalities are one half of a scaling relation, i.e., they show that some relation between exponents which is conjectured to be an equality is at least an inequality in one direction. Further inequalities between critical exponents are in Newman's contribution to [64].

It is somewhat silly to list the conjectures discussed here as open problems. Rather, they form the core of scaling theory and *proving of power laws, calculation of critical exponents and proving universality and the scaling relations, is a whole program.* Only modest beginnings in this direction have been made and the program is likely to keep statistical mechanicians and probabilists busy for a long time.

2.3. *Heuristics for Section* 2.2. For the interested reader we insert here a heuristic section which attempts to make some of the principal conjectures of scaling theory for percolation plausible. More about this can be found in Amit [7] and Fisher [38]. These references also explain how scaling and the renormalization group arose in other physics models. *All further sections are independent of this section, which should be skipped by anyone interested in the flavor of the subject only.*

Power laws and universality seem to have been conjectured purely phenomenologically, on the basis of experimental evidence concerning critical phenomena. In some models such results can be demonstrated, albeit not always rigorously. Also simulations for percolation support these conjectures. The favorite explanation for such conjectures at present rests on renormalization group arguments. A version for percolation might run as follows. Consider first the relation

$$P_{\mathrm{cr}}\{\mathrm{Rad}(W) \geq n\} \approx n^{-1/\delta_r}$$

[given in (1.24)], which only concerns the system at p_c. At criticality there is no favorite length scale. One therefore believes that (in some sense) the "connectivity picture" in $[-n, n]^d$ when suitably scaled looks the same for all large n. In particular, for fixed $\lambda > 1$,

$$\text{(2.12)} \qquad \frac{P_{\mathrm{cr}}\{\mathrm{Rad}(W) \geq \lambda n\}}{P_{\mathrm{cr}}\{\mathrm{Rad}(W) \geq n\}} = P_{\mathrm{cr}}\{\mathrm{Rad}(W) \geq \lambda n | \mathrm{Rad}(W) \geq n\}$$

should have a limit as $n \to \infty$. Indeed, the right-hand side of (2.12) is the probability that some of the endpoints of open paths from $\mathbf{0}$ to $\partial([-n, n]^d)$ can be continued by open paths to $\partial([-\lambda n, \lambda n]^d)$ across $[-\lambda n, \lambda n]^d \backslash [-n, n]^d$. If the limit of (2.12) indeed exists for all $\lambda > 1$, then (1.24) follows; in fact, $P_{\mathrm{cr}}\{\mathrm{Rad}(W) \geq n\}$ even has to be regularly varying (cf. [37], Section 8.8). Similar arguments can be adduced for (2.7). (1.19) seems less intuitive to us, but if one believes that $\#W$ (= "volume" of W) behaves like a power of Rad(W) when the latter is large (for $d = 2$ this is justified in Kesten's contribution to [64]), then (1.24) implies (1.19).

Next we discuss the power laws, which involve powers of $(p - p_c)$, beginning with (2.6). Suppose that we start with site percolation on $\mathscr{G}$ and that for some fixed L we divide space into the cubes

$$(2.13)\quad [j_1L, (j_1+1)L) \times [j_2L, (j_2+1)L) \times \cdots \times [j_dL, (j_d+1)L),$$

$j_i \in \mathbb{Z}$, and identify the cube in (2.13) with the site $(j_1, \dots, j_d)$ of a new copy of $\mathbb{Z}^d$. Call such a cube an (L-) renormalized site. We would like to find a percolation problem for the renormalized sites which is (more or less) equivalent to our original percolation problem. Before attempting this we illustrate the effect of renormalization in a much simpler situation. Let $\{X(v): v \in \mathbb{Z}^d\}$ be an i.i.d. family of random variables with distribution F of mean μ and variance 1. We view μ as the parameter which plays the role of p. The variables $X(v)$ do not necessarily take two values only, but we shall say that the system percolates if

$$\sum_{|v| \le n} X(v) \to \infty, \qquad n \to \infty.$$

By the law of large numbers $\mu_c = 0$ and the system percolates (does not percolate) w.p. 1 if $\mu > \mu_c$ ($\mu < \mu_c$). By the central limit theorem we have to look at the sum of $X(v)$ over a cube of edge size at least $|\mu|^{-2/d}$ to note that $\mu \neq \mu_c = 0$. Thus, the correlation length can be taken to be $|\mu|^{-2/d}$, which is the power law (2.6) with exponent $2/d$ for this example. (Note the "universality"; this statement does not depend on the shape of F.) Renormalization assigns to the L-renormalized site $\mathbf{j} = (j_1, \dots, j_d)$ the random variable

$$Y_L(j) := \frac{1}{L^{d/2}} \sum_{v \text{ in the cube } (2.13)} X(v),$$

whose distribution we denote by $F_L(\cdot)$. F_L is a functional of F, so that we can write $F_L = S(L, F)$ for some transformation S of distribution functions. Note that $S(L_2, S(L_1, F)) = S(L_2L_1, F)$. The original problem for the X's is equivalent to a problem of exactly the same kind for the Y_L's. We have merely replaced F by F_L. Note that for large L, F_L is more or less normal with mean $\mu L^{d/2}$ and variance 1. If F has mean 0, then so does $S(L, F)$, and $S(L, F)$ converges to a standard normal distribution as $L \to \infty$. This limit distribution is independent of the shape of F, and it is a fixed point of the transformation S.

We return to our original percolation problem. In analogy to our example we want to obtain a percolation problem for the renormalized sites. To do this we must decide when a renormalized site is to be called occupied. There is no obvious way to do this, but presumably, if we call the cube in (2.13) occupied when it contains sufficiently many occupied paths on $\mathscr{G}$ (say an occupied path between each of the d pairs of opposite faces), then an infinite cluster of occupied renormalized sites is almost the same as an infinite occupied cluster in the original percolation problem on $\mathscr{G}$. This is only approximately true, because the occupied crossings on $\mathscr{G}$ in two adjacent occupied renormalized sites do not necessarily connect, as illustrated in Figure 3. The assumption is that for $p > p_c$ and large L the probability of a connection is high and that this is a good

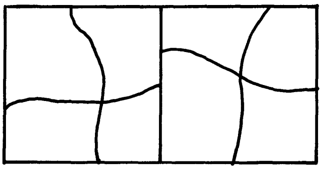

FIG. 3. *Two adjacent occupied renormalized sites. The occupied paths in the left and right square do not necessarily connect.*

approximation. Once this is accepted we can calculate

$$S(L, p) := P_p\{\text{an } L\text{-renormalized site is occupied}\}.$$

The above approximation would say that the probability of an occupied connection on $\mathscr{G}$ over distance n, with parameter p, should behave like the probability of an occupied connection over n/L renormalized sites, i.e., connection over distance n/L in site percolation on $\mathbb{Z}^d$ with parameter $S(L, p)$. In particular, we can expect in some approximate sense

$$\xi(p) = L\xi(S(L, p)). \tag{2.14}$$

For $p < p_c$ this approximation is not convincing. It becomes slightly more acceptable that two adjacent occupied sites are connected, if we define $\mathbf{j}$ to be occupied if there exists an occupied crossing in the i-direction of

$$[j_1L, (j_1 + 1)L] \times \cdots \times [(j_i - 1)L, (j_i + 2)L] \times \cdots \times [j_dL, (j_d + 1)L],$$

for $i = 1, \dots, d$. For $d = 2$ two adjacent occupied sites now necessarily connect (see Figure 4). The new percolation problem is, however, no longer a Bernoulli

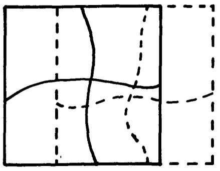

FIG. 4. *Two adjacent renormalized sites which are occupied according to the second definition. The solid and dashed squares are of size* $3L \times 3L$; *the solid and dashed paths are occupied and must connect.*

percolation problem. Sites $\mathbf{j}$ and $\mathbf{k}$ are dependent if $|\mathbf{j} - \mathbf{k}| \le 3$. We are therefore forced to consider p not as a single parameter, but as some infinite-dimensional vector which can describe also the dependence structure of the percolation problem; $S(L, \cdot)$ is then a transformation on this infinite-dimensional parameter space.

If we start with the percolation problem on L_1-renormalized sites (instead of the original problem on $\mathscr{G}$) and combine L_2^d L_1-renormalized sites into one L_2L_1-renormalized site, then the same argument gives

$$\xi(S(L_1, p)) = L_2\xi(S(L_2L_1, p)).$$

In this last step the original problem on $\mathcal{G}$ does not enter, so that S should be (at least in some asymptotic sense) independent of $\mathcal{G}$, which argues for universality, as we shall see. If the original problem is critical, i.e., on the boundary between the regimes with only finite clusters and with an infinite cluster, then the same should hold for the system of renormalized sites. We postulate that if we start at p_c, iteration of $S = S(L, \cdot)$ will take us to some fixed point $\bar{p}$ of S, which does not depend on $\mathcal{G}$ (except through its dimension) or on L, i.e.,

$$S^j(L, p_c) \to \bar{p}, \qquad j \to \infty. \tag{2.15}$$

(Compare this with the behavior of the F_L's in the simple example of i.i.d. random variables, considered above.) We also assume that S can be linearized near $\bar{p}$, so that

$$S(L, p) - \bar{p} \sim \Lambda_L(p - \bar{p}),$$

for some linear operator Λ_L. Furthermore, we assume that Λ_L has a dominant eigenvalue λ_L so that

$$\log\|\Lambda_L^k(\pi)\| - \log\|\pi\| \sim k \log|\lambda_L|,$$

for some norm $\|\cdot\|$, k large and vectors π which are not orthogonal to the eigenspace corresponding to λ_L. All this is somewhat vague, since we have not specified the space of p's on which we are working (and, what is more, do not know how to do this); see [38], Section 5, for a slightly different formulation. We thus hope that

$$\log\|S^k(L, p) - S^k(L, p + \pi)\| - \log\|\pi\| \sim k \log|\lambda_L|, \tag{2.16}$$

for p close to $\bar{p}$ and suitably small π and large k. We now want to start with Bernoulli percolation on $\mathcal{G}$ with parameter p_0 close to $p_c = p_c(\mathcal{G})$, and we want to apply (2.16) with $p = S^j(p_c)$, $p + \pi = S^j(p_0)$. This is unreasonable for $j = 0$, since p_c is not necessarily close to $\bar{p}$, but by (2.15), we can choose $j_0 = j_0(p_c)$ so that $S^{j_0}(p_c)$ is close to $\bar{p}$, and then (2.16) should hold for $p = S^j(p_c)$, $p + \pi = S^j(p_0)$, when $j = j_0$, p_0 close to p_c and k such that $S^{j+k}(p_0)$ is still close to $\bar{p}$. The closer p_0 is to p_c, the larger we should be able to take k (see Figure 5). We expect $|\lambda_L| > 1$, since $|\lambda_L| < 1$ would imply that all systems near p_c would converge to $\bar{p}$ under iterations of S, and all such systems would have similar behavior. This would contradict the fact that arbitrarily close to p_c there are both percolating and nonpercolating systems, by definition of p_c. (We ignore the boundary case $|\lambda_L| = 1$.) Finally, if we start out in the one-dimensional subspace of Bernoulli percolation, with p_0 and p_c numbers close together, we should be able to replace $\|S^{j_0}(p_0) - S^{j_0}(p_c)\|$ by $C|p_0 - p_c|$ with $C = (d/dp)S^{j_0}(L, p)$ evaluated at p_c (provided this derivative exists). Putting all this together we expect that we can let $p_0 \to p_c$ and $k \to \infty$ such that

$$\log\left\|S^{k+j_0}(p_0) - S^{k+j_0}(p_c)\right\| \sim \log C|p_0 - p_c| + k \log|\lambda_L|$$

is bounded away from 0 and ∞. This would give

$$-\log|p_0 - p_c| \sim k \log|\lambda_L|.$$

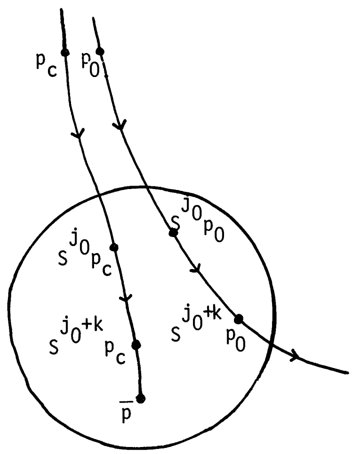

FIG. 5. *The flow of S^jp. (2.16) should work well in the circled neighborhood.*

In addition,

$$\left|\log \xi\left(S^{k+j_0}(p_0)\right)\right| = \left|\log \xi\left(S^{k+j_0}(p_c) + S^{k+j_0}(p_0) - S^{k+j_0}(p_c)\right)\right|$$

can then be expected to stay bounded [since $S^{k+j_0}(p_c) \to \bar{p}$], so that $(k + j_0)$ iterations of (2.14) finally give

$$\log \xi(p_0) \sim k \log L \sim -\frac{\log L}{\log|\lambda_L|}\log|p_0 - p_c|,$$

which is (2.6) with p replaced by p_0 and $\nu = (\log|\lambda_L|)^{-1}\log L$. [Note that $|\lambda_L| > 1$ and hence $\nu > 0$ is consistent with (1.18).]

Once we accept that the correlation length obeys a power law, the remaining argument is that the correlation length is the fundamental length scale, and that dependence on p should be governed only by this length scale. This leads to the scaling ansatz

$$\tau(p, \mathbf{0}, v) \sim |v|^{2-d-\eta} g\left(\frac{|v|}{\xi(p)}\right) \tag{2.17}$$

(cf. Essam [36], Section 4.6) and

$$P_p\{\#W = n\} \sim n^{1-\tau} f_{\pm}\left(\frac{n}{\xi(p)}\right), \tag{2.18}$$

where the $+$ and $-$ signs correspond to $p > p_c$ and $p < p_c$, respectively (cf. [89], Section 3.1.1, or [36], Section 4, for a similar hypothesis; see also [38], Section 5, for further justification). Here g and $f_{\pm}$ are "nice" functions, τ some constant, and the asymptotic relations are supposed to hold when $p \to p_c$, $|v| \to \infty$ $[n \to \infty]$ in such a way that $|v|/\xi(p)$ [respectively, $n/\xi(p)$] tends to a limit. Note that, in view of (2.7), (2.17) states that $\tau(p, \mathbf{0}, v)$ is approximately equal to

$\tau(p_c, \mathbf{0}, v)$ times a correction factor which depends only on $|v|/\xi(p)$ [i.e., $|v|$ measured on the scale $\xi(p)$]. Since $\tau(p, \mathbf{0}, v) \to \tau(p_c, \mathbf{0}, v)$ and $\xi(p) \to \infty$ as $p \to p_c$, we expect that

$$0 < g(0) = \lim_{x \downarrow 0} g(x) < \infty. \tag{2.19}$$

Also, since $\tau(p, \mathbf{0}, v) \to 0$ as $|v| \to \infty$, we should have

$$g(x) \to 0, \quad \text{as } x \to \infty \tag{2.20}$$

[and even exponentially fast, by (1.17)]. (2.19) and (2.20) should also hold with g replaced by $f_{\pm}$. Thus, the scaling hypotheses (2.17) and (2.18) go hand-in-hand with the intuitive interpretation of the correlation length after (1.17).

From (2.6), (2.17) and (2.18) one can obtain the remaining power laws, as well as the scaling relations, by simple manipulations and a few intuitive hypotheses. For example, by (2.18) and (1.7)

$$\chi(p) = \sum_1^\infty n P_p\{\#W = n\}$$

should behave like

$$\sum n^{2-\tau} f\left(\frac{n}{\xi(p)}\right) \sim \int_0^\infty x^{2-\tau} f\left(\frac{x}{\xi(p)}\right) dx$$
$$= [\xi(p)]^{3-\tau} \int_0^\infty y^{2-\tau} f(y)\, dy.$$

Not only does this, together with (2.6), imply (2.5), but it even says that $\gamma = \nu(3 - \tau)$. By similar manipulations (cf. [36], Sections 4.3 and 4.6, and [89], Sections 3.1.2 and 4.2.1 and Appendix 1), one can express also the other exponents in terms of ν and τ. Elimination of τ then yields the proposed scaling relations.

As a final bit of heuristics we point out that (2.17)–(2.19) and the analogue of (2.19) for f indicate that in blocks of size $\xi(p)$ or smaller the "connectivity picture" should have a very similar distribution under P_p as under P_{cr}. I.e., on the scale of the correlation length, the picture is essentially the same as that of a critical system. It is difficult to formulate this in a precise way, but a modest beginning in this direction for $d = 2$ is in [66], where it is proven that

$$0 < C_1 \le \frac{P_p\{\text{Rad}(W) \ge n\}}{P_{\text{cr}}\{\text{Rad}(W) \ge n\}} \le C_2 < \infty, \quad \text{uniformly for } n \le \xi(p).$$

2.4. *Continuum percolation.* For many phenomena it seems unreasonable to impose a lattice structure on the model. For instance, the early work of Flory [39] mentioned previously dealt with gelation. When long polymer molecules in solution form sufficiently many bonds between the molecules, then an infinite connected cluster arises or a gel is formed (see Figure 6). Flory used a branching process as a model for this situation (ignoring cycles of intermolecular bonds), but I do not know of a really satisfactory model for this situation. In any case, it points to the need for models in which space is continuous rather than discrete.

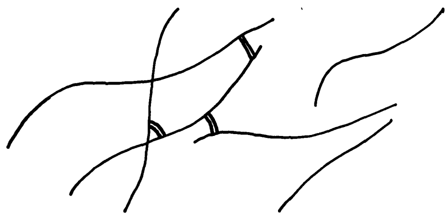

FIG. 6. *The long strands are polymer molecules. Intermolecular bonds (as indicated by the double lines) may form.*

One appealing continuum model (which seems of no use for the gelation phenomena, though) is the following. Let $\mathscr{P} = \{p_i\}_{i \geq 1}$ be the points of a Poisson process in $\mathbb{R}^d$ with constant density λ. Let $\{F_i\}_{i \geq 1}$ be a sequence of i.i.d. figures, independent of $\mathscr{P}$. E.g., F_i might be a ball with center at the origin and (random) radius r_i. Let W be the component of $\mathbf{0}$ of $\bigcup_{i=1}^{\infty}(p_i + F_i)$. $p_i + F_i$ is the figure F_i translated by p_i, $W = \varnothing$ if $\mathbf{0} \notin \bigcup(p_i + F_i)$. If F_i is the ball with center $\mathbf{0}$ and radius r_i, then $\bigcup(p_i + F_i)$ is the union of a countable number of balls, centered at the random points p_i. $\bigcup(p_i + F_i)$ takes the place of the collection of open bonds in our first model. Percolation now corresponds to W being infinite, and accordingly we define the *critical density* as

$$\lambda_c := \sup\{\lambda\colon P_\lambda\{W \text{ is infinite}\} = 0\},$$

where P_λ is the measure governing the model when $\mathscr{P}$ has density λ. There seems to be some ambiguity in this definition, since "W is infinite" may mean "W has infinite Euclidean volume" as well as "W contains infinitely many p_i." Fortunately, these two interpretations lead to the same λ_c (Roy [83]). Questions quite analogous to the ones for ordinary bond percolation can be phrased in this model and some results including estimates of λ_c can be found in Menshikov, Molchanov and Sidorenko [74], Molchanov and Stepanov [75], Hall [48], [49] and Roy [83]. Criteria for percolation of more general (dependent) random fields are also considered in [74] and [75].

One original motivation for this model is given by Gilbert in [42]. Take $d = 2$ and think of p_i as the location of an FM relay station. Its signal can be transmitted up to its horizon, i.e., it can be received by points in the disc

$$D(p_i, 2r_i) := \{x\colon \|x - p_i\|_2 < 2r_i\}.$$

Is the probability that a signal from the station nearest to the origin can be transmitted infinitely far strictly positive? For $r_i = r$, a constant, this is equivalent to the question whether

$$P_\lambda\{\bigcup(p_i + D(\mathbf{0}, r)) \text{ is infinite}\} > 0,$$

since transmission through the stations at $p_{i(1)}, p_{i(2)}, \dots$ is possible only if $\|p_{i(k+1)} - p_{i(k)}\|_2 < 2r$. Actually, a more important type of question is, what the probability is of transmitting a signal across the United States. This leads to the problem of estimating probabilities of open crossings of rectangles; for bond percolation on $\mathbb{Z}^d$ we shall discuss this problem in Section 2.5.

Alternatively to introducing a critical λ in the last model, one may, when $F_i = B(\mathbf{0}, r) := \{x \in \mathbb{R}^d: \|x\|_2 < r\}$ (the ball of radius r), fix λ, as 1 say, and introduce the critical radius

$$r_c := \sup\{r: P\{\text{component of } \mathbf{0} \text{ in } \bigcup(p_i + B(\mathbf{0}, r)) \text{ is infinite}\} = 0\}.$$

This is actually the way Gilbert looked at the problem, and it is also the way in which it appears most naturally in the cluster analysis of Hartigan [55]. The problem in the last paper is statistical. If $X_1, X_2, \dots, X_n$ are i.i.d. observations from a continuous density $f(\cdot)$ on $\mathbb{R}^d$, one wishes to determine how many components the "high-density set" $S(c) := \{x \in \mathbb{R}^d: f(x) \geq c\}$ (for some given $c > 0$) has, by counting the number of r-clusters in the sample which contain at least a fraction $\alpha > 0$ of the observations. Here α and r are at our disposal and an r-cluster of the observations is a maximal subset $\mathscr{C}$ of the observations such that if X_i and X_j belong to $\mathscr{C}$, then there is a chain $X_{i(0)} = X_i, X_{i(1)}, \dots, X_{i(\rho)} = X_j$ of observations in $\mathscr{C}$ from X_i to X_j with $\|X_{i(s+1)} - X_{i(s)}\|_2 \leq r$, $0 \leq s < \rho$. [55] proves that if $S(c)$ consists of $\nu < \infty$ compact components $S_1, \dots, S_\nu$, any pair of which is separated from each other by the set $\{x: f(x) \leq \lambda c\}$ for a certain constant $\lambda = \lambda(c) < 1$, then there exist $r_n = r_n(c)$, $\alpha_n = \alpha_n(f, c)$, $\varepsilon = \varepsilon(f, c) > 0$ and $\varepsilon_n = \varepsilon_n(f, c) \downarrow 0$ such that for $n \to \infty$,

$$\text{(2.21)}\quad \begin{aligned} &P\{\exists\ \nu\ r_n\text{-clusters } \mathscr{C}_1, \dots, \mathscr{C}_\nu \text{ as above, such that } d(S_i, \mathscr{C}_i) \leq \varepsilon_n, \\ &\text{but } \inf\{\|x - y\|_2: x \in \mathscr{C}_i,\ y \in \mathscr{C}_j,\ i \neq j\} > \varepsilon\} \to 1. \end{aligned}$$

Here

$$d(A, B) := \sup_{x \in A} \inf_{y \in B} \|x - y\|_2.$$

It should be intuitively clear that the proof of this consistency result relies on results for the continuum percolation model of this section, since the r_n-clusters are essentially the same as the components of our set $\bigcup(p_i + F_i)$ for $F_i = B(\mathbf{0}, r/2)$. Hartigan [55], page 391, also formulates a conjecture for continuum percolation, which is an analogue of (1.10). If true, this would allow us to take $\lambda = \lambda_n \uparrow 1$ in the conditions for (2.21), and thus would show that components of $S(c)$ can usually be discovered. The Hartigan conjecture has been proven for $d = 2$ ([83]), but not yet for $d \geq 3$. However, the results of [74] Section 8, and [4] leave little doubt that Hartigan's conjecture (in slightly modified form) is true for all $d \geq 2$.

Ramey [80] used minimal spanning trees for another version of Hartigan's cluster analysis. Again properties of continuum percolation are used to derive properties of minimal spanning trees for $X_1, \dots, X_n$ when $d = 2$ (cf. [80], Lemma 3.3.1).

2.5. *Resistance and flow problems*; *crossing probabilities of blocks and reliability.* An old problem in physics, going back more than a century to Maxwell [73], Volume 1, page 365, and Rayleigh [81] is the determination of the effective conductivity of a mixture of materials with known conductivity. For instance, if one mixes a fraction p of a good conductor A with conductivity σ_A and a fraction $1 - p$ of a bad conductor B with corresponding conductivity $\sigma_B < \sigma_A$, how well does the mixture conduct electricity? The answer depends on how the materials are mixed, and is not determined by the numbers p, σ_A, σ_B alone. Stochastic continuum versions of this problem have been treated by Golden, Papanicolaou and Varadhan ([43] and [78]). Here I want to discuss a discrete version. Consider the edges of $\mathbb{Z}^d$ as i.i.d. random resistances with distribution F (concentrated on $[0, \infty]$). Künnemann [69] has already studied this model when each resistance lies in a fixed interval $[a, b]$ with $0 < a < b < \infty$, and also [43] and [78] correspond to this case. To make contact with percolation theory, we consider the extreme case where F has an atom of size p at 1 (ohm) and an atom of size $q = 1 - p$ at ∞ (ohm). The edges with infinite resistance do not conduct electricity, so that we may as well remove them. Thus, we are dealing with the resistance of the random network obtained by removing a fraction q of the edges of $\mathbb{Z}^d$ (and giving each remaining edge the fixed resistance of 1 ohm). Let

$$S_0 := [0, n]^{d-1} \times \{0\} \quad \text{and} \quad S_n := [0, n]^{d-1} \times \{n\} \tag{2.22}$$

be the "bottom" and "top" of the large cube $[0, n]^d$ and for the random network let

$$R_n := \text{resistance in } [0, n]^d \text{ between } S_0 \text{ and } S_n.$$

(For this we have to think of all vertices in one face S_i as connected by superconducting material; see [59], Chapter 11, for a precise definition.) For p small we expect that there are no conducting paths between S_0 and S_n, at least with high probability when n is large. This would result in $R_n = \infty$. On the other hand, for $p = 1$, one easily calculates that for the full network $R_n = n(n+1)^{1-d}$. Thus, one can hope that for large p,

$$r_\infty(p) := \lim_{n\to\infty} n^{d-2} R_n \text{ exists (in some sense) and lies in } (0, \infty). \tag{2.23}$$

In fact, one believes that these two are the only possible asymptotic behaviors, and that the separation between the low and high p regime is again at $p_c(\mathbb{Z}^d, \text{bond})$, the critical probability for bond percolation on $\mathbb{Z}^d$. For $d = 2$ we have the following result, which is a good step in the desired direction (cf. [59], Theorem 11.2, and [15]).

THEOREM. *For* $d = 2$

$$P_p\{R_n = \infty \text{ eventually}\} = 1, \quad \text{if } p < \tfrac{1}{2}, \tag{2.24}$$

$$P_{1/2}\{\lim R_n = \infty\} = 1, \quad \text{if } p = \tfrac{1}{2}, \tag{2.25}$$

and for some constants $0 < C_i$, $\delta < \infty$ *and* θ *defined by* (1.4) *for bond percolation on* $\mathbb{Z}^2$,

$$P_p\left\{0 < C_1[\theta(p)]^{-2} \le \liminf R_n \le \limsup R_n \le C_2\left(p - \tfrac{1}{2}\right)^{-\delta}\right\} = 1, \quad \text{if } p > \tfrac{1}{2}. \tag{2.26}$$

For $d > 2$ we have much less information. (2.24) still holds for $p < p_c(\mathbb{Z}^d, \text{bond})$ [by [59], equation (11.20), and $p_T = p_c$ in [1]]. Also, by [15] [this time with $\theta(\cdot)$

the percolation probability for bond percolation on $\mathbb{Z}^d$; cf. (1.4)]

(2.27) $$P_p\{0 < C_1[\theta(p)]^{-2} \le \liminf n^{d-2}R_n\} = 1, \quad \text{if } p > p_c(\mathbb{Z}^d, \text{bond}).$$

However, an analogue of the other inequality in (2.26) is only known for $p > \hat{p}_c^{\infty}$, where

(2.28) $\hat{p}_c^{[k]} = \hat{p}_c^{[k]}(d) :=$ critical probability for bond percolation on the "slab" $(\mathbb{Z}^+)^2 \times \{0, 1, \dots, k\}^{d-2}$

and

(2.29) $$\hat{p}_c^{\infty} := \lim_{k \to \infty} \hat{p}_c^{[k]}.$$

Then (cf. [15], Theorem 3.6)

(2.30) $$P_p\{\limsup n^{d-2}R_n < \infty\} = 1, \quad \text{if } p > \hat{p}_c^{\infty}.$$

Note that the existence of $r_\infty(p)$ in (2.23) has not even been shown for $d = 2$ (even though the limit does exist in the L^2-sense in the more restricted cases of [43], [69] and [78]). Nevertheless, it is believed to exist for all $p > p_c = p_c(\mathbb{Z}^d, \text{bond})$ and to satisfy a power law

(2.31) $$r_\infty(p) \approx (p - p_c)^{-t}, \qquad p \downarrow p_c.$$

(2.26) together with the bound $\theta(p) \ge C_1(p - p_c)^{\beta_1}$ for $p > p_c$ ([59], Theorem 8.1) sandwiches $r_\infty(p)$ between two powers of $(p - p_c)$ if $d = 2$ (more information about these powers follows from [15] and [16]). This discussion raises the immediate problems:

(2.32P) Is $\hat{p}_c^{\infty}(d) = p_c(\mathbb{Z}^d, \text{bond})$?

Trivially, $\hat{p}_c^{\infty}(d) \ge p_c(\mathbb{Z}^d, \text{bond})$ and equality is believed by many people. An affirmative answer to (2.32) would solve several problems, since there are a number of properties which we believe true above p_c, but can only prove above $\hat{p}_c^{\infty}$.

(2.33P) Prove the existence of $r_\infty(p)$ in some sense for $p > p_c(\mathbb{Z}^d, \text{bond})$.

(2.34P) Establish the power law (2.31) and relate the exponent t to other critical exponents.

In connection with the last problem, it should be pointed out that the fine structure of the resistance network for p close to p_c is still not well understood and it is not yet clear whether t is an "independent" exponent or expressible in terms of other exponents (see [68] and [15] for recent surveys of some work in this direction and also the next section). So far the rigorous bounds in (2.24)–(2.30) and [15] are obtained by estimating

(2.35) $\nu(n) :=$ maximal number of edge-disjoint paths in $[0, n]^d$ from S_0 to S_n.

As pointed out $R_n = \infty$ if $\nu(n) = 0$, whence (2.24) and its analogue for $d > 2$. On the other hand, it is intuitively clear that large $\nu(n)$ should lead to small R_n. To make this quantitative assume that there exist $\nu = \nu(n)$ edge-disjoint paths from S_0 to S_n of respective lengths $l_1, \dots, l_\nu$. A self-avoiding conducting path of length, l, i.e., consisting of l edges, has resistance l ohm, since it consists of l one-ohm resistances in series. Edge-disjoint paths from S_0 to S_n act as resistances in parallel so that familiar rules for combining parallel resistances give

$$R_n \leq \left\{ \sum_1^\nu \frac{1}{l_i} \right\}^{-1} \leq \frac{1}{\nu^2} \sum_1^\nu l_i$$

(2.36) (by harmonic-arithmetic mean inequality)

$$\leq \left(\frac{1}{\nu(n)} \right)^2 dn^d.$$

In the last inequality we used the facts that our paths are edge-disjoint, and that the number of edges in $[0, n]^d$ is dn^d. Thus, (2.30) is reduced to showing

$$\liminf \frac{\nu(n)}{n^{d-1}} > 0, \tag{2.37}$$

and an upper bound for $\limsup n^{d-2} R_n$ will be d times the (-2)th power of the left-hand side of (2.37). It is not clear how good this upper bound is, but in any case the resistance problem seems closely tied to estimating the distribution of the number of edge-disjoint crossings of $[0, n]^d$ from S_0 to S_n. This is a special case of the *flow problem*, to which we return in Section 3.3.

Through considerations of crossing probabilities, percolation theory also makes contact with *reliability theory*. It is not hard to think of the sites (or bonds) of $\mathbb{Z}^d$ as machine parts or switches which can be operating or defective, and the total machine works only if there is a path in $[0, n]^d$ from S_0 to S_n containing only operating parts. The probability of the existence of such a crossing from S_0 to S_n is then the probability that the machine is in working order. The number of disjoint paths with operating parts only is a measure of the reliability of the machine. The more such paths exist, the less likely it is for the machine to break down. Less obvious are the applications of Greene and El Gammal [45] to the manufacture of integrated circuits. The circuit is made up of "processors." Some processors have a defect. One now produces spare rows and columns of processors and after testing of the processors "on chip switches are set to connect nondefective elements into the desired configurations." Identify the processors with the sites of $\mathbb{Z}^2$ and assume they are independently defective (nondefective) with probability q (respectively, p). The connections between nondefective processors are to run in so-called channels (the regions between the rows and columns of the array). One wants the connections to be short, say of length $\leq d$, and not too many connections in one channel, say at most t connections in one channel. [45] discusses problems of the following form:

(a) How to choose connections such that at least Rn^2 nondefective processors from $[0, n]^2$ form a linear chain (not passing through any processor twice) with connections between successive processors in the chain of length $\leq d$ and at most two connections per channel. Here $R \in (0,1)$ is a given number which we would like to be large. The law of large numbers shows that the probability that this can be done tends to 0 as $n \to \infty$ for $R > p$. [45] shows a construction, which is successful with a probability tending to 1 (at a rate which is estimated) when $R < p$ and d is large enough. It divides the array into square blocks. "Good" blocks are those with at least four nondefective processors. With suitable choices the good blocks form a large percolating cluster of blocks, which allows for efficient connections between the blocks (see Figure 7, which is based on [45]).

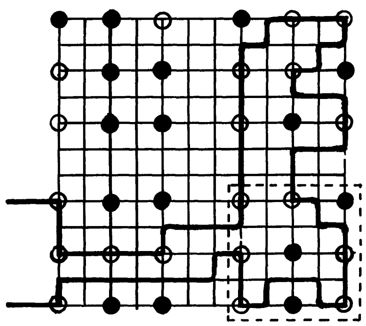

FIG. 7. *A section of an array connected into a chain.* ○ = *nondefective and* ● = *defective processor. Each block contains nine elements. Channels for one wire are provided between elements within a block and for two wires between blocks.*

(b) Find a wiring scheme for the connection of a $k \times k$ array of nondefective processors, from $[0, n]^2$ [or even $[0, R(1 + \varepsilon)n] \times [0, n]$] with $k \geq Rn$ (see Figure 8). It is shown in [45] that for $R < p$ this can be done with a probability

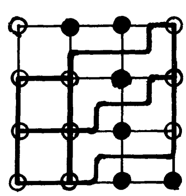

FIG. 8. *Connection of a* 3×3 *nondefective array from a* 4×4 *array.*

tending to 1 ($n \to \infty$) in such a way that the connections have length $O(\sqrt{\log n})$ and so that the number of connections per channel is bounded.

[45] also shows that its requirements on the connection length d are in some sense the best possible ones.

2.6. *The incipient infinite cluster and the ant in the labyrinth.* The behavior of the resistance R_n of the last section, and its critical exponent t, are perhaps the least understood of all the quantities mentioned so far. Clearly, to obtain a power law for $r_\infty(p)$, one needs an understanding of the long open (or conducting) paths when p is close to p_c. This leads one to study the structure of the infinite cluster when $p - p_c$ is very small, but strictly positive. Ideally, one would like to "take the limit" and study the infinite cluster at p_c (the so-called *incipient infinite cluster*). Unfortunately, we know for $d = 2$—and suspect that this is true for all $d \geq 2$—that there exists no infinite cluster at p_c (i.e., with P_{cr}—probability one). How should one then define an object which takes the place of an infinite cluster at criticality? One can try to force the existence of an infinite cluster by means of conditioning. In [62] it is proved for $d = 2$ that there is a probability measure ν on the configuration space (which in our case is compact, being the product of two-point spaces) such that the weak limits

$$(2.38) \qquad \lim_{n \to \infty} P_{cr}\{\cdot \mid \mathbf{0} \to \partial([-n, n]^2)\} \quad \text{and} \quad \lim_{p \downarrow p_c} P_p\{\cdot \mid W \text{ is infinite}\}$$

exist, and both equal $\nu(\cdot)$. Still for $d = 2$, there exists a unique infinite cluster $\tilde{W}$ a.e. $[\nu]$, and $\tilde{W}$ contains the origin. We proposed in [62] to call $\tilde{W}$ the incipient infinite cluster. An obvious question is:

(2.39P) Do the limits in (2.38), with $[-n, n]^d$ in place of $[-n, n]^2$, exist and are they equal, when $d > 2$?

One's first reaction is that one should be able to answer (2.39) affirmatively by simple monotonicity properties or inequalities, but this has not worked so far, so that at present the above approach is restricted to $d = 2$.

Chayes and Chayes [13], Section 7.1, have suggested the eventually invaded region in invasion percolation (see the next section) for the incipient infinite cluster. The advantage of this definition is that it works for all $d \geq 2$, but it is not clear how the definition relates to the one via the limit ν of (2.38). On the other hand, the second limit in (2.38) corresponds intuitively to what one takes as incipient infinite cluster in simulation studies. E.g., in simulating a random walk on the incipient infinite cluster, as discussed below, one often seems to choose a large number, say N, and then simulates N steps of the random walk while simultaneously building the random cluster. One only keeps those trials in which the cluster building does not stop before N steps of the random walk. This amounts to a conditioning similar in spirit to the first limit in (2.38); one conditions on the "cluster being larger than the span of the walks" (cf. [8]). In the simulations of [56] (cf. page L270) the conditioning is a bit more complex. One conditions on the cluster of $\mathbf{0}$ being the largest cluster in some square (and on the frequency of open sites in this square being within 0.0005 of p).

Still another approach to incipient infinite clusters is in [17].

What can one say about the structure of the incipient infinite cluster, and how does it help us with the resistance problem? For $d = 2$ and $\tilde{W}$ the unique infinite cluster chosen according to the measure ν of (2.38), it has been shown that $\tilde{W}$ has asymptotic density $P_{cr}\{\text{Rad}(W) \geq n\}$, in the sense that

$$\nu\left\{\varepsilon \leq \frac{\#(\tilde{W} \cap [-n, n]^2)}{n^2 P_{cr}\{\text{Rad}(W) \geq n\}} \leq \varepsilon^{-1}\right\} \to 1,$$

as $\varepsilon \downarrow 0$, uniformly in n (cf. [62], Theorem 8). Little else is known about the structure of $\tilde{W}$. Much effort has been devoted to the so-called link, nodes and blob picture of the incipient infinite cluster (cf. Stanley and Coniglio in [25], [68], Section 5, and Stanley in [88]). So far this has not yielded a rigorous quantitative relation with the resistance problem.

De Gennes [23] suggested that one should be able to obtain useful information for the resistance problem from what he called "an ant in a labyrinth" (see also the survey of Mitescu and Roussenq in [25]). The labyrinth here is the network of open edges in an incipient infinite cluster for bond percolation. (Similar versions for site percolation exist.) The ant performs a random walk in the labyrinth. Of the several versions considered, the simplest one, to us, is the one in which the ant moves from its position X_n at time n, to its position X_{n+1} at time $(n+1)$ by moving along an open edge incident to X_n; this edge is chosen uniformly from all the open edges incident to X_n. $\{X_n\}$ is really a random walk in a random environment (environment = labyrinth); it is only conditionally Markovian, when the labyrinth is fixed. One studies power laws for

$$E\{\|X_n\|_2^2\} \quad \text{or} \quad P\{X_n = \mathbf{0} | X_0 = \mathbf{0}\}$$

and tries to relate the corresponding exponents to the resistance problem and the δ and δ_r of (1.19) and (1.24). Even though there are some relations known between random walks and resistances (cf. [26]), the only rigorous result for this problem seems to be that if $d = 2$, and the labyrinth is the unique infinite cluster $\tilde{W}$ distributed according to ν, then $\{n^{1/2-\varepsilon}\|X_n\|_2\colon n \geq 1\}$ is a tight family for some $\varepsilon > 0$. Thus, X_n shows *subdiffusive behavior* (cf. [63]). But it would be desirable to solve the following problem:

(2.40P) Find the proper scaling and limit law for X_n.

As a closing problem of this section, which may have a bearing on the structure of the incipient infinite cluster as well as the resistance problem, we mention the length of the shortest open crossing of $[0, n]^d$ from S_0 to S_n [cf. (2.22) for notation], conditional on the existence of such a crossing.

(2.41P) How does the length of the shortest crossing from S_0 to S_n behave under P_{cr}? In particular, does it grow faster than $n^{1+\varepsilon}$ for some $\varepsilon > 0$?

The second question can be paraphrased as, "Are open crossings at criticality necessarily very tortuous?" The numerical evidence of [79] and [44] seems to be not fully conclusive, according to [32]; see also Stanley's paper in [88]. For a related problem see (3.18P).

2.7. *Invasion percolation.* All the preceding models were static. Time played no role and the whole configuration of open and closed edges is chosen at one single moment. A more dynamical model was introduced by de Gennes and Guyon [24], modified by Lenormand and Bories [71] and Chandler, Koplik, Lerman and Willemsen [12], and studied further in [98], [99] and [19]–[21]. The original motivation was to describe the displacement of one fluid by another. Interest in this arises from methods which attempt to recover oil by pumping water into the ground. In this model one assigns to each edge e of $\mathbb{Z}^d$ a value $p(e) \geq 0$. Without attempting to describe the physics accurately, we think of e as a capillary, and $p(e)$ as the minimal pressure which the water must have to force the oil out of this capillary. If water is pumped in only at $\mathbf{0}$, then nothing happens until the pressure reaches

$$\min\{p(e) \colon e \text{ incident to } \mathbf{0}\}. \tag{2.42}$$

Assume that there exists a unique edge e_1 incident to $\mathbf{0}$ for which the minimum in (2.42) is taken on, and set $\hat{W}(1) = \{e_1\}$. If the pressure is increased, oil is first forced out of $W(1)$, and nothing else happens until the pressure reaches

$$\min\{p(e) \colon e \text{ touches } \hat{W}(1), \text{but } e \notin \hat{W}(1)\}. \tag{2.43}$$

After that the oil is forced out of an edge e_2 for which the minimum in (2.43) is achieved. Inductively, $\hat{W}(n)$ will be a connected set of n edges, which contains $\mathbf{0}$; it is called the *invaded region at time n*. $\hat{W}(n+1) = \hat{W}(n) \cup e_{n+1}$, where

$$p(e_{n+1}) = \min\{p(e) \colon e \text{ touches } \hat{W}(n), \text{but } e \notin \hat{W}(n)\}. \tag{2.44}$$

So far we have assumed that the oil can always be forced out, but in practice it is important to take the phenomenon of *trapping* into account. A region R becomes trapped by $\hat{W}(n)$ if R is separated from ∞ by $\hat{W}(n)$. Once R is trapped no oil from R can be displaced, so in the version which takes trapping into account the edges of R become forbidden for invasion when R gets first trapped. (2.44) has to be modified accordingly: The minimum in (2.44) is now only over edges which have not yet been trapped by time n, and e_{n+1} has to be one of those edges.

The model becomes stochastic when one assumes that $\{p(e) \colon e \in \mathbb{Z}^d\}$ is an i.i.d. family of random variables. For simplicity, we take the common distribution of the $p(e)$ to be the uniform distribution on $[0,1]$. [This is not much of a restriction for our purposes, since for any other continuous distribution function F, we can realize $p(e)$ as $F^{-1}(\xi(e))$ for a uniform i.i.d. family $\{\xi(e)\}$, and the invaded region for the p's will be the same as for the ξ's.] The resulting models are called *invasion percolation* (without trapping) and *invasion percolation* (with trapping). Perhaps the most immediate problem is:

> What is the volume fraction of the trapped region in invasion percolation with trapping? More precisely, what is
>
> $$\lim_{N\to\infty} \frac{1}{d(2N)^d} \times \left\{\textit{number of trapped edges in } [-N, N]^d \setminus \bigcup_k \hat{W}(k)\right\} \tag{2.45P}$$
>
> and what is the behavior of the number of edges trapped by $\hat{W}(n)$?

It is useful to allow n to be a random time in the last question. For instance, in simulations one often takes for n the first time when each edge in $[-N, N]^d$ is either trapped by or belongs to $\hat{W}(n)$.

Interesting questions also came up in simulation studies of invasion percolation (without trapping). [98] studied the empirical distribution functions

$$Q_n(x) := \frac{1}{n} \sum_{i=1}^{n} I[p(e_i) \le x]$$

of the invaded edges. [Actually, [98] considered some numerical version of the density of $Q_n(x)$, and even that only at certain stopping times, but it is simpler to discuss Q_n itself.] It was discovered numerically that $[p_c = p_c(\mathbb{Z}^d, \text{bond})]$

$$Q_n(x) \to \begin{cases} x/p_c, & \text{if } x \le p_c, \\ 1, & \text{if } x > p_c. \end{cases} \tag{2.46}$$

One can understand (2.46) intuitively if one knows the following standard method to simulate ordinary bond percolation: Pick the $\{p(e)\colon e \in \mathbb{Z}^d\}$ i.i.d. uniformly distributed on $[0, 1]$, and call an edge e p-open if and only if $p(e) \le p$. It is easy to see that the collection of p-open edges then has the same distribution as the collection of open edges under P_p in bond percolation on $\mathbb{Z}^d$. This method which seems to have been introduced by Hammersley [51], Scheme C, therefore allows us to construct a sample of the open configuration under P_p for all $p \in [0, 1]$ simultaneously. To connect this with (2.46) note that for each $p_0 > p_c$, there is a unique infinite cluster $W(p_0)$ of p_0-open edges. Once the invaded region reaches $W(p_0)$, i.e., once $e_n \subset W(p_0)$ for some n [and hence $p(e_n) \le p_0$], all future invaded edges e_k with $k \ge n$ will belong to $W(p_0)$ [since the invasion can simply proceed picking edges from $W(p_0)$ and have corresponding pressure $\le p_0$]. Thus, one expects for any $p_0 > p_c$ that $p(e_n) \le p_0$ eventually, and that the edges $e_1, e_2, \dots$ picked in the invasion process behave as if they were picked from $W(p_c)$. This would immediately explain the second line in (2.46). But also the first line of (2.46) becomes reasonable; it says that the invaded edges behave asymptotically as if they are samples from the conditional distribution of $p(e)$, given $p(e) \le p_c$. Presumably, the fact that e_n is invaded does not tell us much more about $p(e_n)$ than that $p(e_n)$ is not much higher than p_c.

An interesting fact about the above construction (which is essentially in [51], page 285) is that the percolation probability

$$\theta(p) = P_p\{W \text{ is infinite}\} = P\{p(e_n) \le p \text{ for all } n \text{ in invasion (without trapping) percolation}\}.$$

Unfortunately, there are difficulties with the above explanation. One has to prove that the invasion reaches $W(p_0)$ for $p_0 > p_c$ w.p. 1. For $d = 2$ this is easy, but for $d \ge 3$ it has not yet been proven. Furthermore, we know that there is no infinite cluster $W(p_c)$ when $d = 2$, and we believe that this is also true for any $d \ge 3$. Nevertheless, the above heuristics explain why the eventually invaded region $\bigcup \hat{W}(n)$ was proposed as the incipient infinite cluster (cf. [13], Section 7.1). It also explains why invasion percolation has features similar to a critical ordinary bond percolation system. This is made more precise in [19], where also a good part of (2.46) is proven rigorously.

Chayes, Chayes and Newman [20] also used invasion percolation to obtain the following bounds for ordinary bond percolation on $\mathbb{Z}^d$:

$$\tau(p, \mathbf{0}, x) \le \exp(-C_1(p)|x|), \tag{2.47}$$
$$P_p\{\#W = n\} \le \exp\left(-\frac{C_2(p)}{\log n} n^{(d-1)/d}\right),$$

for p sufficiently large. E.g., (2.47) holds for $p > \hat{p}_c^\infty$ [cf. (1.15) and (2.29) for notation]. These results gain in interest when they are compared with (1.17) (for $p < p_c$), and the lower bounds of [3] and [59] mentioned in connection with (1.26).

[98] and [99] also try to set up a scaling theory for invasion percolation, and, in particular, to interpret critical exponents related to the rate of convergence in (2.46) to critical exponents for Bernoulli percolation. Since we do not yet rigorously have power laws for Bernoulli percolation such relations are somewhat speculative. In the case of invasion percolation with trapping there is even some debate as to whether the analogue of (2.46) requires a new critical probability instead of p_c (cf. [98] and [21]).

3. First-passage percolation.

3.1. *The classical model. The asymptotic shape result.* Originally introduced by Hammersley and Welsh [52] as a generalization of Bernoulli percolation, first-passage percolation has sufficiently distinct problems from Bernoulli percolation to deserve a separate chapter. One now assigns to each edge e of $\mathbb{Z}^d$ a nonnegative random variable $t(e)$. (Site versions of the model—with the t's assigned to sites rather than bonds—have also been considered, but we shall not discuss those here.) $t(e)$ is usually interpreted as the passage time of the edge e, but other interpretations will be discussed below. Throughout we shall make the following assumption:

$$\{t(e) \colon e \in \mathbb{Z}^d\} \text{ are i.i.d.} \tag{3.1}$$

The common distribution function of the $t(e)$ will be denoted by F. The *passage time* of a path r on $\mathbb{Z}^d$ which runs successively through the edges $e_1, \dots, e_n$ is

$$T(r) := \sum_{i=1}^{n} t(e_i).$$

The e_i do not have to be distinct here, but note that loops in a path can only add to its passage time. The *travel time* from u to v is defined as

$$T(u, v) := \inf\{T(r) \colon r \text{ a path from } u \text{ to } v\}.$$

Finally,

$$\tilde{B}(t) = \{v \in \mathbb{Z}^d \colon T(\mathbf{0}, v) \le t\}$$

is the set of vertices which can be reached from the origin by time t. The principal object of study in first-passage percolation is the set $\tilde{B}(t)$ and various of its asymptotic (for large t) properties. This is closely related to problems of finding optimal routes, since $v \in \tilde{B}(t)$ if and only if the fastest route from $\mathbf{0}$ to v has travel time less than or equal to t. To see the relationship with Bernoulli percolation consider the case in which

$$p = P\{t(e) = 1\} = 1 - P\{t(e) = 0\}, \tag{3.2}$$

and identify open edges with those edges of zero-passage time. Then the open cluster W of (1.1) is just the set $\tilde{B}(0)$. Thus, Bernoulli percolation deals with very special aspects of $\tilde{B}$.

To formulate the principal result of first-passage percolation "smooth out" the lattice structure by replacing $\tilde{B}(t)$ by

$$B(t) = \{v + \overline{U}: v \in \tilde{B}(t)\},$$

where

$$\overline{U} = \{x = (x(1), \dots, x(d)): |x(i)| \le \tfrac{1}{2}, 1 \le i \le d\}$$

(the closed unit cube centered at $\mathbf{0}$). The principal result says that $B(t)$ grows linearly with t [possibly at an infinite rate, see (3.6)] and has an asymptotic shape which is not random. The first version of this was proved by Richardson [82]; the following final form is due to Cox and Durrett [22].

(3.3) THEOREM. *Assume that*

$$E \min\{t_1^d, \dots, t_{2d}^d\} < \infty, \tag{3.4}$$

where $t_1, \dots, t_{2d}$ are i.i.d. random variables with distribution F. Then there exists a nonrandom convex set $B_0 \subset \mathbb{R}^d$ with nonempty interior, and which is either compact or equals all of $\mathbb{R}^d$, and has the following property:

(3.5) *If B_0 is compact, then for all $\varepsilon > 0$, $(1 - \varepsilon)B_0 \subset \frac{1}{t}B(t) \subset (1 + \varepsilon)B_0$ eventually w.p. 1.*

(3.6) *If $B_0 = \mathbb{R}^d$, then for all $\varepsilon > 0$, $\{x: |x| \le \varepsilon^{-1}\} \subset \frac{1}{t}B(t)$ eventually w.p. 1.*

If (3.4) fails, then

$$\limsup_{v \to \infty} \frac{1}{|v|} T(\underline{0}, v) = \infty \quad w.p.\ 1. \tag{3.7}$$

The most obvious problem is:

(3.8P) Determine B_0 as a function of F (and the dimension d).

By symmetry B_0 must be invariant under permutations of the coordinates and reflections in the coordinate hyperplanes. If the rightmost point of B_0 on the first coordinate axis is $(\mu^{-1}, 0, \dots, 0)$, then by convexity B_0 lies between the "diamond" $\{x: \Sigma_1^d |x(i)| \le \mu^{-1}\}$ and the cube $\{x: |x| = \max|x(i)| \le \mu^{-1}\}$ (see Figure 9). Durrett and Liggett [29] showed in a remarkable two-dimensional

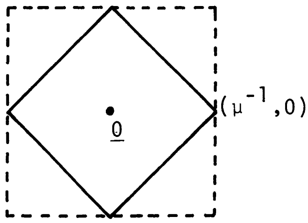

FIG. 9. B_0 *must contain the diamond* (*with the solidly drawn boundary*) *and be contained in the cube* (*with dashed boundary*).

example that the boundary of B_0 can contain straight-line segments, but one has not been able to calculate B_0 entirely in a single example (excluding the trivial case when F assigns unit mass to one point). The only general theorem about B_0 is the distinction between the cases (3.5) and (3.6) (cf. [61], Theorem 1.15, and [13], Theorem 4.4).

THEOREM. $B_0 = \mathbb{R}^d$ *if and only if*

$$F(0) \geq p_c(\mathbb{Z}^d, \textit{bond}). \tag{3.9}$$

We should also mention that [despite (3.7)] without *any* moment condition on F, one can replace $T(\mathbf{0}, v)$ by a quantity $\hat{T}(\mathbf{0}, v)$ such that

$$\frac{1}{|v|}|T(\mathbf{0}, v) - \hat{T}(\mathbf{0}\, v)| \to 0 \quad \text{in probability,}$$

as $|v| \to \infty$ and such that (3.5) and (3.6) hold with $B(t)$ replaced by

$$\hat{B}(t) := \{v + \overline{U}\colon \hat{T}(\mathbf{0}, v) \leq t\}.$$

This was proven first for $d = 2$ in [22], and later for $d \geq 3$ in [61], Theorem 3.1.

Of course, one has done simulation studies of B_0. The earliest ones may be the ones of Eden [31] in connection with *Eden's growth model.* In this model an animal grows in discrete time. Its state at time n, A_n, is a connected set of $(n + 1)$ vertices of $\mathbb{Z}^d$. At time 0, we start with $A_0 = \{\mathbf{0}\}$. A_{n+1} is formed from A_n by adding at random one of the sites adjacent to A_n, but not in A_n. If the probability of adding a site v to A_n to obtain A_{n+1} is proportional to the number of neighbors of v in A_n (see Figure 10), then one obtains a model which is equivalent (in a sense to be explained) to the study of $B(t)$ when F is the exponential distribution,

$$F(x) = (1 - e^{-x})^+. \tag{3.10}$$

A second model picks all sites adjacent to A_n with the same probability for being added to form A_{n+1}; in this model all $\times$'s in Figure 10 have probability 1/7 of being added. This version corresponds to the site version of first-passage percolation, still with the F of (3.10). The relation between Eden's model and first-passage percolation is that [under (3.10)] A_n has exactly the same

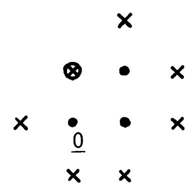

FIG. 10. *The solid circles denote A_2. One of the sites marked $\times$ will be added to form A_3. In the first model the site marked $\otimes$ is twice as likely to be added as the other $\times$'s (hence with probability* 2/8).

distribution as $\tilde{B}(t_n)$, where

$$t_n = \inf\{t: \tilde{B}(t) \text{ contains at least } (n+1) \text{ vertices}\}.$$

Eden [31] already suggested that one investigate the asymptotic shape of A_n, but no progress was made until Richardson [82] realized the above equivalence, which, together with (3.5), implies that A_n (when suitably normalized) has the asymptotic shape B_0 corresponding to (3.10).

Most simulations seem to have been done for the second model (the site version) in two dimensions. Here the Eden animal looks very circular for large n. Nevertheless, it is *not* believed that B_0 is a Euclidean ball when F is given by (3.10). For one thing, we know that it is not a Euclidean ball in either the bond or site version when d is large (cf. [61], Section 8, and Dhar's contribution to [88]) and any reasons for spherical symmetry in low dimensions should also apply in high dimensions. Also recent simulations suggest that even for $d = 2$, B_0 is not a (Euclidean) disc (cf. [100]).

3.2. *First-passage times and the time constant.* The full theorem about the existence of B_0 was not proven right at the start of the subject. Rather one started looking at the "rightmost" point of B_0. More precisely, Hammersley and Welsh introduced

$$\begin{aligned} a_{0,n} &:= T(\mathbf{0},(n,0,\dots,0)) \quad \text{and} \\ b_{0,n} &:= T(\mathbf{0},\{x: x(1) = n\}), \end{aligned} \tag{3.11}$$

the so-called *point-to-point* and *point-to-line* (or *rather point-to-hyperplane*) *passage times*. They proved [under the assumption $Et(e) < \infty$] that $n^{-1}a_{0,n} \to \mu$ in probability for some constant $\mu = \mu(F, d)$, the so-called *time constant*. They also proved similar convergence results for related passage times and conjectured that also $n^{-1}b_{0,n} \to \mu$ in probability. It was not until Kingman proved his famous subadditive ergodic theorem (cf. [72], Theorem 6.2.6 for a recent version) that one could obtain the following stronger result: If

$$E\{\min\{t_1,\dots,t_{2d}\}\} < \infty \tag{3.12}$$

[with t_i as in (3.4)], then

$$\lim_{n\to\infty} \frac{1}{n} a_{0,n} = \lim_{n\to\infty} \frac{1}{n} b_{0,n} = \mu \quad \text{w.p. 1 and in } L^1 \tag{3.13}$$

(see [87] for this and a good survey of the early results in the subject). Again one can drop the moment condition (3.12) if one replaces T by the $\hat{T}$ [mentioned after (3.9)] in the definitions (3.11); in this case one obtains for the corresponding $\hat{a}_{0,n}$ only convergence w.p. 1 (and not in L^1) in (3.13). (3.13) is contained in Theorem 3.3, since $a_{0,n} \le t$ is equivalent to $(n, 0, \dots, 0) \in B(t)$ and similarly for $b_{0,n}$. In fact, the intersection of the boundary of B_0 with the positive first coordinate axis must be the point $(1/\mu, 0, \dots, 0)$. In practice, one first proves (3.13) and uses it to obtain Theorem 3.3.

A subproblem of (3.8) is to find μ as a functional of F (and d). This and (3.8) seem rather hopeless at the moment, since μ is only obtained by Kingman's subadditive ergodic theorem. This theorem almost never leads to a computable limit. Except for the easy estimate

$$0 \le \mu \le E\{t(e)\} = \int x \, dF(x), \tag{3.14}$$

we do not even have good methods to estimate μ. Note that $\mu = 0$ corresponds to (3.6) and is equivalent to (3.9). The second inequality in (3.14) is strict unless F is concentrated on one point (cf. [52], Theorem 4.1.9).

If $t(e)$ has the Bernoulli distribution of (3.2), then μ becomes a function of p, with $\mu = 0$ if and only if $1 - p \ge p_c$ [by (3.9)]. It is believed that for $1 - p < p_c$, μ has a power law

$$\mu \approx \left(p_c - (1 - p)\right)^{\theta}.$$

In fact, it is known [16] that μ behaves essentially like $[\xi(1 - p)]^{-1}$ [cf. (1.16)] so that $\theta = \nu$ provided ν and θ exist [cf. (2.6)]. (Note that our p here corresponds to $1 - p$ in [16].)

The next most obvious question is:

(3.15P) Find a limit law for $\dfrac{1}{\gamma(n)}(\theta_{0,n} - n\mu)$ or $\dfrac{1}{\gamma(n)}(\theta_{0,n} - \delta(n))$ for suitable constants $\gamma(n) \to \infty$ and $\delta(n)$.

Here, and in the rest of this section, θ stands for a or b. Extremely little progress has been made with this problem. Some (poor) estimates on the rate of convergence of $n^{-1}\theta_{0,n}$ to μ have been proven in [61], Section 5. The proof there shows that one would first like to know:

(3.16P) At what rate does $n^{-1}E\theta_{0,n} \to \mu$?

For $d = 2$ and F the Bernoulli distribution of (3.2) with $p = \frac{1}{2} = p_c(\mathbb{Z}^2, \text{bond})$, i.e., at the critical point, [16] proves the remarkable result that $E\theta_{0,n}$ lies between two positive multiples of $\log n$ so that $\{(\log n)^{-1}\theta_{0,n}\}$ is a tight family. The arguments of [86] and [84] already proved in this critical case that $P\{\theta_{0,n} \le \varepsilon \log n\}$ is small for small ε, uniformly in n, so that any limit distribution of $(\log n)^{-1}\theta_{0,n}$ has no mass at zero.

Already Hammersley and Welsh [52] pointed out that (3.16) could be approached via the so called *height problem*. Call a path r on $\mathbb{Z}^2$ from $\mathbf{0}$ to

$(n, 0, \dots, 0)$ [or $\{x(1) = n\}$] a *route* for $a_{0,n}$ (or $b_{0,n}$, respectively) if $T(r) = a_{0,n}$ [$T(r) = b_{0,n}$]. It is known ([87], Section 4.3) that such routes exist when $d = 2$, but not (yet?) known when $d > 2$. We therefore restrict ourselves now to $d = 2$. Define

$$M(r) = \max\{x(2): x = (x(1), x(2)) \text{ a point of } r\}.$$

(3.17P) Does the maximal height $\frac{1}{n}\max\{M(r): r \text{ a route for } \theta_{0,n}\} \to 0$?

See [61], Section 9, and [52], Section 8, for more details. We note that the answer to (3.17) is negative for some lattices other than $\mathbb{Z}^2$ and some F. In fact, the above-mentioned example of [29] in which B_0 has a flat edge gives, after rotation over 45°, an example where (3.17) fails on $\mathbb{Z}^2$ rotated over 45° (compare also [61], Lemma 9.10).

Another problem basically going back to [52] concerns the length of routes.

(3.18P) Does $\lim \frac{1}{n}\min\{\text{number of edges in } r: r \text{ a route for } \theta_{0,n}\}$ exist?

Results on (3.18) can be found in [87], Chapter 8, [58] and [101]. This problem is related to (2.41P).

Further problems are listed in Chapter 9 of [61].

3.3. *Flow problems and a higher-dimensional generalization of first-passage percolation.* Most of this section is taken from [60]. So far we interpreted the $t(e)$ as the passage time of e. We now wish to view it as a *capacity*, i.e., as the maximal amount of fluid which can pass through e per unit time. We study the *maximal flow* from the bottom to the top of the box

$$B = B(n_1, \dots, n_{d-1}, m) := [0, n_1] \times \cdots \times [0, n_{d-1}] \times [0, m],$$

subject to these capacity restrictions. This is defined as follows. Write

$$S(0) = [0, n_1] \times \cdots \times [0, n_{d-1}] \times \{0\} \quad \text{and}$$
$$S(m) = [0, n_1] \times \cdots \times [0, n_{d-1}] \times \{m\},$$

for the "bottom" and "top" of B. A permissible flow from $S(0)$ to $S(m)$ in B is an assignment of a number $0 \le f(e) \le t(e)$ and a direction to each edge e in B, which satisfies for each vertex $v \notin S(0) \cup S(m)$,

$$\Sigma^+ f(e) - \Sigma^- f(e) = 0, \tag{3.19}$$

where Σ^+ [Σ^-] is the sum over all edges e in B incident to v and directed towards v (away from v). We think of $f(e)$ as the amount of fluid flowing through e per unit time in the direction assigned to e. (3.19) says that the total inflow at v equals the total outflow at v. The total flow from $S(0)$ to $S(m)$ is

$$\Sigma_0^- f(e) - \Sigma_0^+ f(e),$$

where Σ_0^+ [Σ_0^-] is the sum over all edges with one endpoint in $S(0)$ and the other endpoint in $B/S(0)$ and directed towards (away from) the endpoint in $S(0)$. The maximal flow from $S(0)$ to $S(m)$ in B is

$$\phi(n_1, n_2, \dots, n_{d-1}, m) := \text{maximum of total flow from } S(0) \text{ to } S(m) \text{ over all permissible flows.}$$

A special case of interest to us is if the distribution F of $t(e)$ is concentrated on $\{0,1\}$ as in (3.2), then one can send one unit of fluid per unit time through a path from $S(0)$ to $S(m)$ all of whose edges have capacity 1 ("open edges") and no fluid through any path containing an edge of capacity 0 ("closed edges"). It is intuitively clear that in this case the maximal flow is equal to the maximal number of edge disjoint open paths in B from $S(0)$ to $S(m)$ (see [9], Theorem 3.5). This is precisely the quantity which we wanted to get a grip on in the resistance problem of Section 2.5.

We return to the case with general $t(e) \geq 0$. One way to study ϕ is by means of the max-flow min-cut theorem of Ford and Fulkerson (e.g., [9], Chapter 3.1), which equates ϕ to the minimal capacity of a cut set. Specifically, for any set E of edges define its value or capacity as

$$V(E) = \sum_{e \in E} t(e).$$

A set E of edges *separates* $S(0)$ *from* $S(m)$ *in* B if there is no path on $\mathbb{Z}^d$ in B from $S(0)$ to $S(m)$ which avoids E. (Thus, $B \setminus E$ is no longer connected.) E is called an $(S(0), S(m))$-*cut* if it is a minimal separating set of this kind. The max-flow min-cut theorem now gives

$$\phi(n_1, n_2, \dots, n_{d-1}, m) = \min\{V(E): E \text{ an } (S(0), S(m))\text{-cut}\}. \tag{3.20}$$

The relation with first-passage percolation appears when we describe the cut sets in the two-dimensional case. Let $\mathscr{L}^*$ be the dual graph of $\mathbb{Z}^2$. We may think of $\mathscr{L}^*$ as $\mathbb{Z}^2 + (\frac{1}{2}, \frac{1}{2})$, the graph $\mathbb{Z}^2$ shifted by $(\frac{1}{2}, \frac{1}{2})$ (see Figure 11). $\mathscr{L}^*$ has a

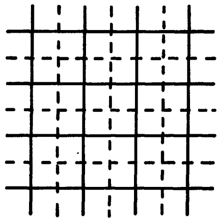

FIG. 11. $\mathbb{Z}^2$ *and* $\mathscr{L}^*$. *The edges of* $\mathbb{Z}^2$ *are drawn as solid segments and those of* $\mathscr{L}^*$ *as dashed segments.*

vertex at the center of each face of $\mathbb{Z}^2$. Each edge e^* of $\mathscr{L}^*$ bisects a unique edge e of $\mathbb{Z}^2$ and vice versa. We call such a pair e^* and e *associated edges*. The edges of $\mathscr{L}^*$ are in a 1–1 correspondence with the edges of $\mathbb{Z}^2$ by this association. The $(S(0), S(m))$-cuts on $\mathbb{Z}^2$ (which separate the bottom edge from the top edge in $[0, n] \times [0, m]$) are now precisely the sets of edges of $\mathbb{Z}^2$ which are associated to the edges of $\mathscr{L}^*$ in a self-avoiding path on $\mathscr{L}^*$ in $[-\frac{1}{2}, n + \frac{1}{2}] \times [\frac{1}{2}, m - \frac{1}{2}]$

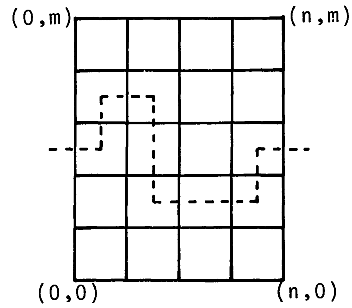

FIG. 12. *The dashed path is an* $(S(0), S(m))$*-cut.*

from $\{-\frac{1}{2}\} \times [\frac{1}{2}, m - \frac{1}{2}]$ to $\{n + \frac{1}{2}\} \times [\frac{1}{2}, m - \frac{1}{2}]$ (see Figure 12). This follows from Whitney's theorem (cf. [87], pages 9–11). It simplifies language if we call a set E^* of edges of $\mathscr{L}^*$ also an $(S(0), S(m))$-cut if the edges of $\mathbb{Z}^2$ associated to edges of E^* form an $(S(0), S(m))$-cut. If, in addition, we set

$$t(e^*) = t(e), \quad \text{if } e^* \text{ and } e \text{ are associated,}$$
$$V(E^*) = \sum_{e^* \in E^*} t(e^*) = \text{value of } E^*,$$

for any set E^* of edges of $\mathscr{L}^*$, then we obtain from (3.20) for $d = 2$,

(3.21) $\phi(n, m) =$ maximal flow from $S(0)$ to $S(m)$ in $[0, n] \times [0, m] =$ $\min\{V(E^*)$: E^* is a self-avoiding path in $[-\frac{1}{2}, n + \frac{1}{2}] \times [\frac{1}{2}, m - \frac{1}{2}]$ from $\{-\frac{1}{2}\} \times [\frac{1}{2}, m - \frac{1}{2}]$ to $\{n + \frac{1}{2}\} \times [\frac{1}{2}, m - \frac{1}{2}]\}$.

If we interpret $t(e^*)$ as the passage time of the edge e^*, rather than its capacity, then for a path r^* on $\mathscr{L}^*$, $V(r^*)$ just becomes what we called its passage time $T(r^*)$ in Section 3.1. It is therefore natural to call the last member of (3.21) the *line-to-line passage time* from $\{-\frac{1}{2}\} \times [\frac{1}{2}, m - \frac{1}{2}]$ to $\{n + \frac{1}{2}\} \times [\frac{1}{2}, m - \frac{1}{2}]$ [compare (3.11)]. Since $\mathscr{L}^*$ is just a shifted copy of $\mathbb{Z}^2$, it comes as no surprise that n^{-1} times this line-to-line passage time also converges w.p. 1 and in L^1 to $\mu(F, 2)$ if the $\{t(e)\colon e \in \mathbb{Z}^2\}$ are i.i.d. with distribution F satisfying $\int t\,dF(t) < \infty$ and if m does not grow too fast. We thus have the following result from [47]:

(3.22) If $Et(e) < \infty$ and $n \to \infty$, $m \to \infty$, in such a way that $\frac{1}{n}\log m \to 0$, then $\frac{1}{n}\phi(n, m) \to \mu(F, 2)$ w.p. 1 and in L^1.

We reach new ground when we try to follow a similar procedure in dimension > 2. (3.20) remains valid, but how to describe the $(S(0), S(m))$-cuts now? To make some use of geometrical intuition *we restrict ourselves to* $d = 3$. As done in [65], we replace $\mathscr{L}^*$ by $\mathbb{Z}^3 + (\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$, and the analogue of a dual edge now becomes a *plaquette*, i.e., a unit square with corners at vertices of $\mathscr{L}^*$. Alternatively, the plaquettes are the faces of the unit cubes with centers at vertices of $\mathbb{Z}^3$. Again each edge e of $\mathbb{Z}^3$ is associated with a unique plaquette π^* of $\mathscr{L}^*$, namely the plaquette π^* which bisects e. For such a pair we again set $t(\pi^*) = t(e)$, and

$$V(E^*) := \sum_{\pi^* \in E^*} t(\pi^*) = \text{value of } E^*,$$

for any set E^* of plaquettes. Again we call a collection of plaquettes an $(S(0), S(m))$-cut if the collection of associated edges on $\mathbb{Z}^3$ forms an $(S(0), S(m))$-cut. Even though we mentally picture the cuts as some kind of surface of plaquettes which runs between the top and bottom of the box $[0, n_1] \times [0, n_2] \times [0, m]$, we have, unfortunately, no good description of the class of cuts. Nevertheless, it is natural to ask for the minimal values of certain surfaces which correspond to the point-to-point, point-to-line and line-to-line passage times. This leads to a generalization of first-passage percolation: Find minimal surfaces rather than minimal paths. One could even go further and ask for minimal l-dimensional objects in d-space, $l < d$. We do not know, however, what the analogue of the asymptotic shape Theorem 3.3 could be.

It is ambiguous what the analogues of the passage times $a_{0,n}$ and $b_{0,n}$ should be. Recall that $a_{0,n}$ is defined as the infimum of $T(r)$ over the collection of paths r whose endpoints are fixed at $\mathbf{0}$ and $(n, 0, \dots, 0)$. Aizenman, Chayes, Chayes, Fröhlich and Russo [2], Section 1(ii), argue that this suggests as an analogue of $a_{0,n}$ the infimum of $V(E^*)$ over a collection of surfaces E^* with fixed boundary. Here the boundary of E^* is defined as

$\partial E^* :=$ collection of edges of $\mathscr{L}^*$ which belong to an odd number of plaquettes in E^*.

Similarly, $b_{0,n}$ was defined as the infimum of $T(r)$ over paths with one endpoint fixed, but the other endpoint only partially restricted, namely to lie in the hyperplane $\{x(1) = n\}$. For reasons explained more fully in [60] and [65] we opted for the following definitions of $\alpha(n_1, n_2)$, $\beta(n_1, n_2)$ as analogues of $a_{0,n}$ and $b_{0,n}$, respectively:

$$B_N = B_N(n_1, n_2) = [0, n_1] \times [0, n_2] \times [-N, N],$$
$$G_{\pm N} = G_{\pm N}(n_1, n_2) = [0, n_1] \times [0, n_2] \times \{\pm N\},$$

(3.23) $\alpha(n_1, n_2) = \inf\{V(E^*)$: E^* is a collection of plaquettes such that ∂E^* consists exactly of the edges in the perimeter of $[-\frac{1}{2}, n_1 + \frac{1}{2}] \times [-\frac{1}{2}, n_2 + \frac{1}{2}] \times \{\frac{1}{2}\}$, and such that E^* separates G_{-N} from G_{+N} in B_N for some $N\}$,

(3.24) $\beta(n_1, n_2) = \inf\{V(E^*)$: E^* is a connected set of plaquettes which contains the point $(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$ and which separates G_{-N} from G_N in B_N for some $N\}$.

One can also define an analogue of the line-to-line passage time, but of more interest is the dual, and equivalent, quantity namely the maximal flow $\phi(n_1, n_2, m)$.

The principal result of [65] proves analogues of (3.13) and (3.22) with the above replacements for $a_{0,n}$, $b_{0,n}$ and the line-to-line passage times. This may be viewed as an indication that (3.23) and (3.24) are the "right" definitions. The results are formulated here as Theorems 3.28 and 3.32. Unfortunately, we can prove Theorem 3.28 only when $F(0)$, the atom of F at the origin, is sufficiently small. Lemma 3.25 serves to define the interval of $F(0)$ values for which Theorem 3.28 works. The hope is that someone will prove this result for all $F(0) < 1 - p_c(\mathbb{Z}^3, \text{bond})$. If that can be done, then one will obtain a fairly complete analogue of the first-passage percolation results for the "minimal surfaces" and maximal flows.

(3.25) LEMMA. *There exists a* $p_0 \geq 1/27$ *with the following property: For every distribution function F with*

$$F(0-) = 0, \qquad F(0) < p_0, \tag{3.26}$$

there exist constants $\Theta = \Theta(F) > 0$, $0 < C_i = C_i(F) < \infty$ *such that*

(3.27) $P\{$*there exists a connected set* E^* *of* n *plaquettes of* $\mathscr{L}^*$ *which contains the point* $(-\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$ *and with* $V(E^*) \leq \Theta n\} \leq C_1 e^{-C_2 n}$, $n \geq 0$.

(3.28) THEOREM. *Assume F satisfies* (3.26) *or more generally* $F(0-) = 0$ *and* (3.27). *If, in addition,*

$$Ee^{\gamma t(\pi^*)} = \int_{[0,\infty)} e^{\gamma x}\, dF(x) < \infty, \tag{3.29}$$

for some $\gamma > 0$, *then there exists a number* $\nu = \nu(F) \leq \int_{[0,\infty)} x\, dF(x) < \infty$ *which is strictly positive and such that*

$$\lim_{n_1, n_2 \to \infty} \frac{1}{n_1 n_2} \alpha(n_1, n_2) = \lim_{n_1, n_2 \to \infty} \frac{1}{n_1 n_2} \beta(n_1, n_2) = \nu \quad w.p.1. \tag{3.30}$$

Moreover, if $n_1, n_2, m \to \infty$ *in such a way that*

$$(\max\{n_1, n_2\})^{-1+\delta} \log m \to 0, \text{ for some } \delta > 0,$$

then also

$$\lim_{n_1, n_2 \to \infty} \frac{1}{n_1 n_2} \phi(n_1, n_2, m) = \nu \quad w.p.1. \tag{3.31}$$

(3.32) THEOREM. *Assume that*

$$F(0-) = 0, \qquad F(0) > 1 - p_c(\mathbb{Z}^3, \text{bond}) \quad \text{and} \quad \int_{[0,\infty)} x^6\, dF(x) < \infty.$$

Then

$$\limsup_{n_1, n_2 \to \infty} \frac{1}{n_1 + n_2} \theta(n_1, n_2) < \infty \quad w.p.1,$$

for $\theta = \alpha$ *or* β. *A fortiori* (3.30) *holds with* $\nu = 0$.

Furthermore, there exists a constant $C_3 = C_3(F) < \infty$ *such that w.p.* 1

$$\phi(n_1, n_2, m) = 0, \quad \textit{for all sufficiently large } n_1, n_2,$$

whenever $m = m(n_1, n_2) \to \infty$ *as* $n_1, n_2 \to \infty$ *in such a way that*

$$\liminf_{n_1, n_2 \to \infty} \frac{m(n_1, n_2)}{\log(n_1 n_2)} \geq C_3.$$

We conclude with the following problem:

(3.33P) Prove (3.30) and (3.31) for all F with $F(0) < 1 - p_c(\mathbb{Z}^3, \text{bond})$. In particular, show that $\nu > 0$ for such F. [One can probably relax (3.29) as well, but that is of less interest.]

As remarked before, the Bernoulli case, when $t(e)$ has the distribution (3.2), is useful for the resistance problem. In particular, (3.31) implies for $d = 3$ for the $\nu(n)$ of (2.35),

$$\lim \frac{1}{n^2} \nu(n) = \nu.$$

A proof of (3.31) with $\nu > 0$ for F determined by (3.2) with $1 - p = F(0) < 1 - p_c(\mathbb{Z}^3, \text{bond})$, would therefore prove that (2.30) holds for all $p > p_c(\mathbb{Z}^3, \text{bond})$ [compare (2.37)].

Acknowledgment. The author is indebted to R. Durrett for various illuminating comments and helpful references, as well as the example discussed below (2.13).

Note added in proof. In a forthcoming paper by J. T. Chayes, L. Chayes, G. R. Grimmett, H. Kesten and R. H. Schonmann it is shown that the limit in (1.16) exists for *all p*. However, it has not been proven that this limit is strictly positive for $p > p_c$.

It has been shown by H. Tasaki (private communication), and independently by J. T. Chayes and L. Chayes in the preprint, "On the upper critical dimension of Bernoulli percolation," that for $d < 6$ it is impossible for all the scaling relations (including hyperscaling) to hold with the mean field values of the exponents. Thus the upper critical dimension of Section 2.2 has to be at least 6.

REFERENCES

[1] Aizenman, M. and Barsky, D. J. (1987). Sharpness of the phase transition in percolation models. *Comm. Math. Phys.* **108** 489–526.

[2] Aizenman, M., Chayes, J. T., Chayes, L., Fröhlich, J. and Russo, L. (1983). On a sharp transition from area law to perimeter law in a system of random surfaces. *Comm. Math. Phys.* **92** 19–69.

[3] Aizenman, M., Delyon, F. and Souillard, B. (1980). Lower bounds on the cluster size distributions. *J. Statist. Phys.* **23** 267–280.

[4] Aizenman, M., Kesten, H. and Newman, C. M. (1987). Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. *Comm. Math. Phys.* (See also contribution to [64].)

[5] AIZENMAN, M. and NEWMAN, C. M. (1984). Tree graph inequalities and critical behavior in percolation models. *J. Statist. Phys.* **36** 107–143.

[6] AIZENMAN, M. and NEWMAN, C. M. (1986). Discontinuity of the percolation density in one dimensional $1/|x-y|^2$ percolation models. *Comm. Math. Phys.* **107** 611–648.

[7] AMIT, D. J. (1978). *Field Theory, the Renormalization Group, and Critical Phenomena.* McGraw-Hill, New York.

[8] BEN-AVRAHAM, D. and HAVLIN, S. (1982). Diffusion on percolation clusters at criticality. *J. Phys. A* **15** L691–L697.

[9] BOLLOBÁS, B. (1979). *Graph Theory: An Introductory Course.* Springer, New York.

[10] BROADBENT, S. R. (1954). In discussion of symposium on Monte Carlo methods. *J. Roy. Statist. Soc. Ser. B* **16** 68.

[11] BROADBENT, S. R. and HAMMERSLEY, J. M. (1957). Percolation processes. I. Crystals and mazes. II. *Proc. Cambridge Philos. Soc.* **53** 629–641 and 642–645.

[12] CHANDLER, R., KOPLIK, J., LERMAN, K. and WILLEMSEN, J. F. (1982). Capillary displacement and percolation in porous media. *J. Fluid Mech.* **119** 249–267.

[13] CHAYES, J. T. and CHAYES, L. (1986a). Percolation and random media. In *Critical Phenomena, Random Systems and Gauge Theories, Les Houches Session XLIII 1984* (K. Osterwalder and R. Stora, eds.) 1000–1142. North-Holland, Amsterdam.

[14] CHAYES, J. T. and CHAYES, L. (1986b). An inequality for the infinite cluster density in percolation. *Phys. Rev. Lett.* **56** 1619–1622. (See also contribution to [64].)

[15] CHAYES, J. T. and CHAYES, L. (1986c). Bulk transport properties and exponent inequalities for random resistor and flow networks. *Comm. Math. Phys.* **105** 133–152.

[16] CHAYES, J. T., CHAYES, L. and DURRETT, R. (1986). Critical behavior of the two-dimensional first passage time. *J. Statist. Phys.* **45** 933–951.

[17] CHAYES, J. T., CHAYES, L. and DURRETT, R. (1987). Inhomogeneous-percolation problems and incipient infinite clusters. *J. Phys. A* **20** 1521–1530.

[18] CHAYES, J. T., CHAYES, L. and FRÖHLICH, J. (1985). The low temperature behavior of disordered magnets. *Comm. Math. Phys.* **100** 399–437.

[19] CHAYES, J. T., CHAYES, L. and NEWMAN, C. M. (1985). The stochastic geometry of invasion percolation. *Comm. Math. Phys.* **101** 383–407.

[20] CHAYES, J. T., CHAYES, L. and NEWMAN, C. M. (1987a). Bernoulli percolation above threshold: An invasion percolation analysis. *Ann. Probab.* **15** 1272–1287.

[21] CHAYES, J. T., CHAYES, L. and NEWMAN, C. M. (1987b). The stochastic geometry of invasion percolation. II. Unpublished manuscript.

[22] COX, J. T. and DURRETT, R. (1981). Some limit theorems for percolation processes with necessary and sufficient conditions. *Ann. Probab.* **9** 583–603.

[23] DE GENNES, P. G. (1976). La percolation: Un concept unificateur. *La Recherche* **7** 919–927.

[24] DE GENNES, P. G. and GUYON, E. (1978). Lois génerales pour l'injection d'un fluide dans un milieu poreux aléatoire. *J. Mécanique* **17** 403–432.

[25] DEUTSCHER, G., ZALLEN, R. and ADLER, J., eds. (1983). *Percolation Structures and Processes. Ann. Israel Phys. Soc.* **5**. Adam Hilger, Bristol.

[26] DOYLE, P. G. and SNELL, J. L. (1984). *Random Walks and Electrical Networks.* Carus Monograph 22. Math. Assoc. Amer., Washington.

[27] DURRETT, R. (1984). Oriented percolation in two dimensions. *Ann. Probab.* **12** 999–1040.

[28] DURRETT, R. (1985). Some general results concerning the critical exponents of percolation processes. *Z. Wahrsch. verw. Gebiete* **69** 421–437.

[29] DURRETT, R. and LIGGETT, T. M. (1981). The shape of the limit set in Richardson's growth model. *Ann. Probab.* **9** 186–193.

[30] DURRETT, R. and NGUYEN, B. (1985). Thermodynamic inequalities for percolation. *Comm. Math. Phys.* **99** 253–269.

[31] EDEN, M. (1961). A two-dimensional growth process. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **4** 223–239. Univ. California Press.

[32] EDWARDS, B. F. and KERSTEIN, A. R. (1985). Is there a lower critical dimension for chemical distance? *J. Phys. A* **18** L1081–L1086.

[33] EFROS, A. (1982a). What is the theory of percolation? *Kvant* **2** 2–9. (In Russian.)

[34] EFROS, A. L. (1982b). *Physics and Geometry of Disorder*. Nauka, Moscow. (In Russian.)

[35] ESSAM, J. W. (1972). Percolation and cluster size. In *Phase Transitions and Critical Phenomena* (C. Domb and M. S. Green, eds.) **2** 197–270. Academic, New York.

[36] ESSAM, J. W. (1980). Percolation theory. *Rep. Progr. Phys.* **43** 833–912.

[37] FELLER, W. (1971). *An Introduction to Probability Theory and Its Applications* **2**, 2nd ed. Wiley, New York.

[38] FISHER, M. E. (1983). Scaling, universality and renormalization group theory. In *Critical Phenomena* (F. J. W. Hahne, ed.). *Lecture Notes in Phys.* **186** 1–138. Springer, New York.

[39] FLORY, P. J. (1941). Molecular size distribution in three dimensional gelation. I–III. *J. Amer. Chem. Soc.* **63** 3083–3100.

[40] FORTUIN, C. M. and KASTELEYN, P. W. (1972). On the random cluster model. *Physica* **57** 536–564.

[41] FRISCH, H. L. and HAMMERSLEY, J. M. (1963). Percolation processes and related topics. *J. Soc. Indust. Appl. Math.* **11** 894–918.

[42] GILBERT, E. N. (1961). Random plane networks. *J. Soc. Indust. Appl. Math.* **9** 533–543.

[43] GOLDEN, K. and PAPANICOLAOU, G. (1982). Bounds for effective parameters of heterogeneous media by analytic continuation. *Comm. Math. Phys.* **90** 473–491.

[44] GRASSBERGER, P. (1985). On the spreading of two-dimensional percolation. *J. Phys. A* **18** L215–L219.

[45] GREENE, J. W. and EL GAMMAL, A. (1984). Configuration of VLSI arrays in the presence of defects. *J. Assoc. Comput. Mach.* **31** 694–717; and J. W. Greene's Ph.D. thesis of same title, Stanford Univ.

[46] GRIMMETT, G. (1976). On the number of clusters in the percolation model. *J. London Math. Soc.* (2) **13** 346–350.

[47] GRIMMETT, G. and KESTEN, H. (1984). First-passage percolation, network flows and electrical resistances. *Z. Wahrsch. verw. Gebiete* **66** 335–366.

[48] HALL, P. (1985). On continuum percolation. *Ann. Probab.* **13** 1250–1266.

[49] HALL, P. (1986). Clump counts in a mosaic. *Ann. Probab.* **14** 424–458.

[50] HAMMERSLEY, J. M. (1959). Bornes supérieures de la probabilité critique dans un processus de filtration. In *Le Calcul des Probabilités et ses Applications* 17–37. CNRS, Paris.

[51] HAMMERSLEY, J. M. (1963). A Monte Carlo solution of percolation in a cubic lattice. In *Methods in Computational Physics* (B. Alder, S. Fernbach and M. Rotenberg, eds.) **1** 281–298. Academic, New York.

[52] HAMMERSLEY, J. M. and WELSH, D. J. A. (1965). First-passage percolation, subadditive processes, stochastic networks and generalized renewal theory. In *Bernoulli, Bayes, Laplace Anniversary Volume* (J. Neyman and L. M. Le Cam, eds.) 61–110. Springer, New York.

[53] HAMMERSLEY, J. M. and WELSH, D. J. A. (1980). Percolation theory and its ramification. *Contemp. Phys.* **21** 593–605.

[54] HARRIS, T. E. (1960). A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.* **56** 13–20.

[55] HARTIGAN, J. A. (1981). Consistency of single linkage for high-density clusters. *J. Amer. Statist. Assoc.* **76** 388–394.

[56] KAPITULNIK, A., AHARONY, A., DEUTSCHER, G. and STAUFFER, D. (1983). Self similarity and correlations in percolation. *J. Phys. A* **16** L269–L274.

[57] KASTELEYN, P. W. and FORTUIN, C. M. (1969). Phase transitions in lattice systems with random local properties. *J. Phys. Soc. Japan* **26** (Suppl.) 11–14.

[58] KESTEN, H. (1980). On the time constant and path length of first-passage percolation. *Adv. in Appl. Probab.* **12** 848–863.

[59] KESTEN, H. (1982). *Percolation Theory for Mathematicians*. Birkhäuser, Boston.

[60] KESTEN, H. (1985). First-passage percolation and a higher dimensional generalization. *Contemp. Math.* **41** 235–251.

[61] KESTEN, H. (1986a). Aspects of first-passage percolation. *Lecture Notes in Math.* **1180** 125–264. Springer, New York.

[62] KESTEN, H. (1986b). The incipient infinite cluster in two-dimensional percolation. *Probab. Theory Related Fields* **73** 369–394.

[63] KESTEN, H. (1986c). Subdiffusive behavior of random walk on a random cluster. *Ann. Inst. H. Poincaré Sect. B (N.S.)* **22** 425–487.

[64] KESTEN, H., ed. (1987a). *Percolation Theory and Ergodic Theory of Infinite Particle Systems.* IMA Volumes in Math. and its Applications **8**. Springer, New York.

[65] KESTEN, H. (1987b). Surfaces with minimal random weights and maximal flows: A higher dimensional version of first-passage percolation. *Illinois J. Math.* **31** 99–166.

[66] KESTEN, H. (1987c). Scaling relations for 2D-percolation. *Comm. Math. Phys.* **108** 109–156.

[67] KESTEN, H. and ZHANG. Y. (1987). Strict inequalities for some critical exponents in two-dimensional percolation. *J. Statist. Phys.* **46** 1031–1055.

[68] KIRKPATRICK, S. (1979). Models of disordered materials, course no. 5. In *La Matière mal condensée/Ill Condensed Matter, Les Houches Session XXXI, 1978* (R. Balian, R. Maynard and G. Toulouse, eds.). North-Holland, Amsterdam.

[69] KÜNNEMANN, R. (1983). The diffusion limit for reversible jump processes on $\mathbb{Z}^d$ with ergodic random bond conductivities. *Comm. Math. Phys.* **90** 27–68. Effective conductivity on a lattice as a limit of box conductivities. Ph.D. thesis, Univ. Heidelberg.

[70] KUULASMAA, K. (1982). The spatial general epidemic and locally dependent random graphs. *J. Appl. Probab.* **19** 745–758.

[71] LENORMAND, R. and BORIES, S. (1980). Description d'un mécanisme de connexion de liaison destiné à l'étude du drainage avec piégeage en milieu poreux. *C. R. Acad. Sci. Paris Sér. B* **291** 279–282.

[72] LIGGETT, T. M. (1985). *Interacting Particle Systems.* Springer, New York.

[73] MAXWELL, J. C. (1873). *A Treatise on Electricity and Magnetism.* Clarendon, Oxford.

[74] MENSHIKOV, M. V., MOLCHANOV, S. A. and SIDORENKO, A. F. (1986). Percolation theory and some applications. *Itogi Nauki i Tekhniki (Series of Probability Theory, Mathematical Statistics, Theoretical Cybernetics)* **24** 53–110.

[75] MOLCHANOV, S. A. and STEPANOV, A. K. (1983). Percolation in random fields. I, II. *Theoret. and Math. Phys.* **55** 478–484 and 592–599.

[76] NGUYEN, B. G. (1987a). Correlation lengths for percolation processes. *J. Statist. Phys.* **46** 517–523.

[77] NGUYEN, B. G. (1987b). Typical cluster size for 2-dim percolation processes. Preprint.

[78] PAPANICOLAOU, G. C. and VARADHAN, S. R. S. (1979). Boundary value problems with rapidly oscillating random coefficients. In *Random Fields, Coll. Math. Soc. János Bolyai* **27**. Esztergom. North-Holland, Amsterdam.

[79] PIKE, R. and STANLEY, H. E. (1981). Order propagation near the percolation threshold. *J. Phys. A* **14** L169–L177.

[80] RAMEY, D. B. (1982). A non-parametric test of bimodality with applications to cluster analysis. Ph.D. thesis, Yale Univ.

[81] LORD RAYLEIGH (1892). On the influence of obstacles arranged in rectangular order upon the properties of a medium. *Philos. Mag.* **54** 481–502.

[82] RICHARDSON, D. (1973). Random growth in a tesselation. *Proc. Cambridge Philos. Soc.* **74** 515–528.

[83] ROY, R. (1987). Ph.D. thesis, Cornell Univ.

[84] RUSSO, L. (1978). A note on percolation. *Z. Wahrsch. verw. Gebiete* **43** 39–48.

[85] RUSSO, L. (1981). On the critical percolation probabilities. *Z. Wahrsch. verw. Gebiete* **56** 229–237.

[86] SEYMOUR, P. D. and WELSH, D. J. A. (1978). Percolation probabilities on the square lattice. *Ann. Discrete Math.* **3** 227–245.

[87] SMYTHE, R. T. and WIERMAN, J. C. (1978). *First-Passage Percolation on the Square Lattice. Lecture Notes in Math.* **671**. Springer, New York.

[88] STANLEY, H. E. and OSTROWSKY, N., eds. (1986). *On Growth and Form.* Martinus Nijhoff, Dordrecht.

[89] STAUFFER, D. (1979). Scaling theory of percolation clusters. *Phys. Rep.* **54** 1–74.

[90] STAUFFER, D. (1981). Scaling properties of percolation clusters. In *Disordered Systems and Localization* (C. Castellani, C. DiCastro and L. Peliti, eds.). *Lecture Notes in Phys.* **149** 9–25. Springer, New York.

[91] STAUFFER, D. (1985). *Introduction to Percolation Theory*. Taylor and Francis, London.

[92] SYKES, M. F. and ESSAM, J. W. (1964). Exact critical percolation probabilities for site and bond problems in two dimensions. *J. Math. Phys.* **5** 1117–1127.

[93] TOULOUSE, G. (1974). Perspectives from the theory of phase transitions. *Nuovo Cimento B* **23** 234–240.

[94] VAN DEN BERG, J. and KESTEN, H. (1985). Inequalities with applications to percolation and reliability. *J. Appl. Probab.* **22** 556–569.

[95] WIERMAN, J. C. (1978). On critical probabilities in percolation theory. *J. Math. Phys.* **19** 1979–1982.

[96] WIERMAN, J. C. (1981). Bond percolation on honeycomb and triangular lattices. *Adv. in Appl. Probab.* **13** 293–313.

[97] WIERMAN, J. C. (1982). Percolation theory. *Ann. Probab.* **10** 509–524.

[98] WILKINSON, D. and WILLEMSEN, J. F. (1983). Invasion percolation: A new form of percolation theory. *J. Phys. A* **16** 3365–3376.

[99] WILLEMSEN, J. F. (1984). Investigations on scaling and hyperscaling for invasion percolation. *Phys. Rev. Lett.* **52** 2197–2200.

[100] ZABOLITZKY, J. G. and STAUFFER, D. (1986). Simulation of large Eden clusters. *Phys. Rev. A* **34** 1523–1530.

[101] ZHANG, Y. and ZHANG, Y. C. (1984). A limit theorem for N_{0n}/n in first-passage percolation. *Ann. Probab.* **12** 1068–1076.

CORNELL UNIVERSITY
DEPARTMENT OF MATHEMATICS
ITHACA, NEW YORK 14853