Lectures in

APPLIED MATHEMATICS

Volume 27

Mathematics of Random Media

Werner E. Kohler
Benjamin S. White
Editors

American Mathematical Society

The Proceedings of the 1989 AMS-SIAM Summer Seminar on Mathematics of Random Media were prepared by the American Mathematical Society with support from the following sources: Air Force Office of Scientific Research, Army Research Office, and the National Science Foundation through NSF Grant DMS-8814807.

1980 *Mathematics Subject Classification* (1985 *Revision*).
Primary: 60H10, 60J60, 60K35, 82A42, 82A43; Secondary: 35R60, 60H25, 60JXX, 82A70.

Library of Congress Cataloging-in-Publication Data

AMS-SIAM Summer Seminar on Mathematics of Random Media (1989: Virginia Polytechnic Institute and State University)

Mathematics of random media/Werner E. Kohler, Benjamin S. White, editors.

p. cm. – (Lectures in applied mathematics, ISSN 0075-8485; v. 27)

Proceedings of the AMS-SIAM Summer Seminar on Mathematics of Random Media were prepared by the American Mathematical Society, held 5/29–6/9/89 at Virginia Polytechnic Institute and State University.

Includes bibliographical references.

ISBN 0-8218-1133-9

1. Stochastic processes–Congresses. 2. Random fields—Congresses. 3. Probabilities—Congresses. 4. Mathematical physics—Congresses. I. Kohler, W. E. (Werner E.), 1939– . II. White, Benjamin S. (Benjamin Steven), 1945– . III. Title. IV. Series: Lectures in applied mathematics (American Mathematical Society); v. 27.

QC20.7.S8A48 1989 90-27442
519.2–dc20 CIP

Printed in the United States of America.

This publication was typeset using $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX, the American Mathematical Society's TEX macro system.

10 9 8 7 6 5 4 3 2 1 94 93 92 91

Contents

Preface ix

The Contact Process, 1974–1989
RICK DURRETT 1

Is the Contact Process Dead?
LAWRENCE F. GRAY 19

Limiting Behavior of a One-Dimensional System with Long Range Interactions
THOMAS M. LIGGETT 31

Topics in Percolation
CHARLES M. NEWMAN 41

The Lace Expansion and the Upper Critical Dimension for Percolation
GORDON SLADE 53

Mandelbrot Percolation in Two and Three Dimensions
GLEN H. SWINDLE 65

Intersection Probabilities for Random Walks
GREGORY F. LAWLER 73

A Summary of Results about Cyclic Cellular Automata
ROBERT FISCH 87

Hydrodynamical Behavior of Asymmetric Attractive Particle Systems One Example: One-Dimensional Nearest-Neighbors Asymmetric Simple Exclusion Process
JEAN-PIERRE FOUQUE 97

Statistical Mechanics Applied to Random-Media Problems
GEORGE STELL 109

A Control Formulation for Constrained Markov Processes
THOMAS G. KURTZ 139

Diffusion Approximation for a Class of Transport Processes with Physical Reflection Boundary Conditions
C. COSTANTINI 151

Random Perturbations of Dynamical Systems: A New Approach
YURI KIFER 163

Partial Differential Systems with a Small Parameter and Diffusion Process with a Discrete Component
A. EIZENBERG AND M. FREIDLIN 175

Asymptotics for the Solution of the Exterior Dirichlet Problem for Second-order Elliptic Operators with Small First-order Perturbations
DIMITRI IOFFE AND ROSS PINSKY 185

Asymptotic Behavior of Invariant Density of Diffusion Process with Small Fluctuation
TOSHIO MIKAMI 193

Lyapunov Exponents for White and Real Noise Driven Two-Dimensional Systems
MARK PINSKY AND VOLKER WIHSTUTZ 201

Asymptotic Methods for Markov Jump Processes
B. J. MATKOWSKY, Z. SCHUSS, AND C. TIER 215

Uniform Solution of Kramers' Problem by a Direct Approach
T. NAEH, M. M. KLOSEK, B. J. MATKOWSKY, AND Z. SCHUSS 241

On Random Differential Delay Equations with a Small Parameter
G. YIN AND K. M. RAMACHANDRAN 269

Homogenization Error Estimates for Random Elliptic Operators
V. YURINSKIĬ 285

The Effective Conductivity of Composites
OSCAR P. BRUNO 293

On a General Concept in Optimal Material Layout
K. A. LURIE 301

Microstructure and Effective Properties of Random Media
S. TORQUATO 323

Classical Transport in Quasiperiodic Media
KENNETH GOLDEN 359

Tuning Up a Quasicrystal
J. D. MAYNARD 375

Localization Estimates for Off-Diagonal Disorder
WILLIAM G. FARIS 391

Order Statistics and Renormalization
WILLIAM G. FARIS 407

Wave and Particle Propagation in One-Dimensional and Quasi-One-Dimensional Random Media
L. A. PASTUR 417

Water Waves in Shallow Channels of Randomly Varying Depth
A. NACHBIN AND G. C. PAPANICOLAOU 429

Approximations by Delta-correlated Random Processes and Diffusive Approximation in Stochastic Problems
V. I. KLYATSKIN 447

Some Aspects of Wave Propagation Theory in Layer Random Media
V. I. GOLAND, V. I. KLYATSKIN, AND I. O. YAROSHCHUK 477

Reflection of Waves by Randomly Layered Media
WERNER KOHLER, GEORGE PAPANICOLAOU, AND BENJAMIN WHITE 487

Preface

In recent years there has been remarkable growth in the mathematics of random media. This collection of papers by leading researchers provides a current overview of this rapidly-developing field. As will be apparent to the reader of this volume, the field has deep scientific and technological roots as well as purely mathematical ones in the theory of stochastic processes.

The papers collected here were presented at the 1989 Summer Seminar in Applied Mathematics, sponsored jointly by the American Mathematical Society and the Society for Industrial and Applied Mathematics, and held at Virginia Polytechnic Institute and State University from May 29–June 9, 1989. In addition to new results on stochastic differential equations and Markov processes, fields whose elegant mathematical techniques are of continuing value in the application areas, the conference was organized around four main themes.

Systems of interacting particles are normally thought of in connection with the fundamental problems of statistical mechanics, but have also been used to model diverse phenomena including computer architectures and the spread of biological populations. They have also enjoyed a more recent whimsical popularity as a computer recreation. Powerful mathematical techniques have been developed for their analysis, and a number of important systems are now well understood.

Random perturbations of dynamical systems have also been used extensively as models in physics, chemistry, biology, and engineering. Among the recent unifying mathematical developments is the theory of large deviations, which enables the accurate calculation of the probabilities of rare events. For these problems approaches based on effective but formal perturbation techniques parallel rigorous mathematical approaches from probability theory and partial differential equations. We include representative papers of forefront research of both types.

Effective medium theory, otherwise known as the mathematical theory of homogenization, consists of techniques for predicting the macroscopic properties of materials from an understanding of their microstructure. This theory is fundamental, for example, in the science of composites, for the theoretical determination of electrical and mechanical properties. Furthermore, the inverse problem is potentially of great technological importance in the design of composite materials which have been optimized for some specific use.

Mathematical theories of the propagation of waves in random media have been used to understand phenomena as diverse as the twinkling of stars, the corruption of data in geophysical exploration, and the quantum mechanics of disordered solids. Especially effective methods now exist for waves in randomly stratified, one-dimensional media. A unifying theme is the mathematical phenomenon of localization, which occurs when a wave propagating into a random medium is attenuated exponentially with propagation distance, with the attenuation caused solely by the mechanism of random multiple scattering.

We hope this Proceedings mirrors the excitement of the conference it records. We thank the other members of the organizing committee, Marty Day, Rick Durrett, and Graeme Milton, for their considerable efforts and expertise. Our special thanks are for conference coordinator Betty Verducci of the AMS, whose cheerful competence actually made administration a pleasure. We were supported by grants from the National Science Foundation, the Army Research Office and the Air Force Office of Scientific Research.

Werner Kohler,
Blacksburg, VA

Benjamin S. White,
Annandale, NJ

Conference Co-Chairmen

Lectures in Applied Mathematics
Volume 27 (1991)

The Contact Process, 1974–1989

RICK DURRETT

Abstract. We describe the current state of knowledge concerning the contact process, which, thanks to recent results of Bezuidenhout and Grimmett, is rather complete.

1. Introduction. The contact process was introduced by Harris [22]. It is a simple model of the spread of a biological population. The state at time t, $\xi_t \subset \mathbf{Z}^d$. The points in ξ_t are thought of as occupied and the system evolves as follows:

(i) if $x \in \xi_t$, then x becomes vacant at rate 1;
(ii) if $x \notin \xi_t$, then x becomes occupied at a rate

$$\lambda|\{y \in \xi_t : |y-x| = 1\}|.$$

Here $|z| = |z_1| + \cdots + |z_d|$ for $z \in \mathbf{Z}^d$ and $|A| =$ the number of points in A if $A \subset \mathbf{Z}^d$. In other words, vacant sites become occupied at a rate proportional to the numer of occupied neighbors.

For later purposes it is useful to have a concrete construction of the process (due to Harris [23]); so we give that now. For each $x \in \mathbf{Z}^d$, let $\{U_n^x : n \geq 1\}$ be the arrival times of a Poisson process with rate 1. As the reader can probably guess from the rate, at time U_n^x we kill the particle at x (if one is present). To generate the births for each $x, y \in \mathbf{Z}^d$ with $|x-y| = 1$, we let $\{T_n^{x,y} : n \geq 1\}$ be the arrival times of a Poisson process with rate λ. At time $T_n^{x,y}$ there is a birth

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60K35.

Partially supported by the National Science Foundation, the Army Research Office through the Mathematical Sciences Institute at Cornell University, and a Guggenheim fellowship.

at y if x is occupied and y is vacant. It is not hard to show that this "graphical representation" can be used to construct on the same space copies of the process starting from any $A \subset \mathbf{Z}^d$. We will give the details later.

The first thing to notice about the contact process is that it "dies out" if $\lambda < 1/(2d)$. Suppose for the sake of drawing pictures that $d = 1$ and consider two things that can happen when $|\xi_t| = 4$:

$$\begin{array}{llllllllllllll} & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ \text{birth rate} & & \lambda & & \lambda+\lambda & & \lambda & & \lambda & & \lambda & \lambda & & \lambda = 8\lambda, \end{array}$$

$$\begin{array}{lllllllll} & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ \text{birth rate} & \lambda & & & & & & \lambda & = 2\lambda. \end{array}$$

A little thought reveals that in $\mathbf{Z}^d$ when $|\xi_t| = k$, the death rate is k and the birth rate is at most $2d\lambda k$, the maximum occurring when no two particles are adjacent. Let ξ_t^0 denote the process with $\xi_0^0 = \{0\}$. Comparing this process with a birth and death process shows

$$P_\lambda(\xi_t^0 \neq \varnothing \text{ for all } t) = 0 \quad \text{for } \lambda \le 1/(2d).$$

The second picture above shows that when $|\xi_t| = k$, the birth rate in $d = 1$ may be as small as 2λ. This fact makes it much more difficult to prove "survival", i.e., $\Omega_\infty \equiv \{\xi_t^0 \neq \varnothing \text{ for all } t\}$ has positive probability. The first and easiest thing to observe is that $P_\lambda(\Omega_\infty)$ is a nondecreasing function of λ; so there is a critical value $\lambda_f = \inf\{\lambda : P_\lambda(\Omega_\infty) > 0\}$. Here the subscript f is for "survival from finite sets" to distinguish it from another critical value that we will define later. Harris [22] gave an argument that shows $\lambda_f \le 1328$ in $d = 1$. His constant is ridiculously large, but try to do better! By using much different methods Holley and Liggett [24] have shown that $\lambda_f \le 2/d$. Numerically $\lambda_f \approx 1.65$ in $d = 1$, and $\lambda_f \approx 0.41$ in $d = 2$ (see Brower, Furman, and Moshe [3] and Grassberger and de la Torre [17]) and it has been shown that $2d\lambda_f \to 1$ as $d \to \infty$. (See Holley and Liggett [25] or Griffeath [20].)

Having shown that $\lambda_f < \infty$ the next question to answer is: What happens when the process survives? Our first step in answering this question is to introduce some general theory. Let ξ_t^A denote the process with $\xi_0^A = A$. The contact process is *attractive*, i.e., if $A \subset B$ then we can construct copies of the process on the same space with $\xi_t^A \subset \xi_t^B$ for all t. Indeed, it is easy to check that the graphical representation introduced above has this property. Let ξ_t^1 denote the process starting from $\xi_t^1 = \mathbf{Z}^d$. An important consequence of being attractive is that as $t \to \infty$, $\xi_t^1 \Rightarrow \xi_\infty^1$. Here $\Rightarrow$ denotes weak convergence, which in this setting is just convergence of the finite-dimensional distributions $P(A \subset \xi_t^1, B \cap \xi_t^1 = \varnothing)$.

The convergence $\xi_t^1 \Rightarrow \xi_\infty^1$ is easy to see intuitively. $\mathbf{Z}^d \supset \xi_s^1$ for $s > 0$ and the system is attractive; so ξ_t^1 is decreasing in time (i.e., $P(A \subset \xi_t^1)$ is decreasing) and hence has a limit. The contact process is a *Feller process*, i.e., if we have a sequence of initial configurations $\xi_0^n \Rightarrow \xi_0^\infty$ then $\xi_t^n \Rightarrow \xi_t^\infty$. By analogy with what happens when we iterate a continuous nondecreasing function from $[0, 1]$ into itself, we should expect ξ_∞^1 to be the largest fixed point (i.e., stationary distribution), and it is. Turning the last intuition into a proof is not difficult or exciting. The reader can find a full account of this and other facts that we use without giving a reference in Liggett [**28**] or Durrett [**6**].

In the last paragraph we have argued that ξ_∞^1 is the largest stationary distribution. There is nothing to guarantee that $\xi_\infty^1 \neq \delta_\varnothing$, the pointmass on the empty configuration, and indeed we will have $\xi_\infty^1 = \delta_\varnothing$ if λ is too small. This motivates the definition of our second critical value $\lambda_e = \inf\{\lambda : \xi_\infty^1 \neq \delta_\varnothing\}$. With two critical values introduced, it is natural if somewhat optimistic to ask if they are equal. This is not true in general (see Durrett and Gray [**8**], [**9**] for an example in $d = 2$) but the answer for the contact process is YES. The key to the proof is *duality*, which we will introduce by giving the promised construction of the process from the graphical representation. Our first step is to write δ's at the points (x, U_n^x) in space time $\mathbf{Z}^d \times [0, \infty)$ and draw arrows from $(x, T_n^{x,y})$ to $(y, T_n^{x,y})$. We say there is a path from $(x, 0)$ to (y, t) if there is a sequence of times $s_0 = 0 < s_1 < s_2 < \cdots < s_{n+1} = t$ and spatial locations $x = x_0, x_1, \ldots, x_n = y$ so that

(i) for $i = 1, 2, \ldots, n$ there is an arrow from (x_{i-1}, s_i) to (x_i, s_i);
(ii) for $i = 0, 1, \ldots, n$ the vertical segments $\{x_i\} \times [s_i, s_{i+1}]$ do not contain any δ's.

To define the contact process starting from the initial configuration A we let

$$\xi_t^A = \{y : \text{ for some } x \in A \text{ there is a path from } (x, 0) \text{ to } (y, t)\}.$$

Since the arrows in a path indicate births and the absence of δ's indicates that the particles did not get killed before they gave birth, it is easy to see that the recipe above gives the contact process.

The nice thing about the last representation is that it allows us to reverse time. We say that there is a path down from (y, t) to $(x, t-s)$

if $(x, t-s)$ can be reached from (y, t) by going down segments with no δ's and across arrows in a direction opposite to their orientation. A little thought shows that if we let

$$\hat{\xi}_s^{(B,t)} = \{x : \text{ for some } y \in B \text{ there is a path down from } (y, t) \text{ to } (x, t-s)\}$$

then $\{\hat{\xi}_s^{(B,t)} : s \le t\} \stackrel{d}{=} \{\xi_s^B : s \le t\}$ and $\{\xi_t^A \cap B \neq \varnothing\} = \{A \cap \hat{\xi}_t^{(B,t)} \neq \varnothing\}$; so

$$P(\xi_t^A \cap B \neq \varnothing) = P(A \cap \xi_t^B \neq \varnothing). \tag{1.1}$$

The last equation says that the contact process is "self-dual". Letting $A = \{0\}$ and $B = \mathbf{Z}^d$, we have $P(\xi_t^0 \neq \varnothing) = P(0 \in \xi_t^1)$. Letting $t \to \infty$, shows

$$P(\xi_t^0 \neq \varnothing \text{ for all } t) = P(0 \in \xi_\infty^1);$$

so $\lambda_e = \lambda_f$. For reasons that will only become clear when the reader considers more general attractive systems, we will use λ_f to denote their common value.

Having identified the two critical values, we are now ready to state our results. The first was an open problem for almost 15 years.

(1.2) THEOREM. $P_{\lambda_f}(\Omega_\infty) = 0$.

As we will explain in Section 2, the proof of (1.2) is done by showing

(*) If $P_\lambda(\Omega_\infty) > 0$, then, when viewed on suitable length and time scales, the contact process dominates oriented percolation.

Once this is done it is routine to use existing technology to show

(1.3) COMPLETE CONVERGENCE THEOREM. *Let* $\tau^A = \inf\{t : \xi_t^A = \varnothing\}$. *If* $P_\lambda(\Omega_\infty) > 0$ *then*

$$\xi_t^A \Rightarrow P(\tau^A < \infty)\delta_\varnothing + P(\tau^A = \infty)\xi_\infty^1 \quad \text{as } t \to \infty.$$

Here we are using ξ_∞^1 to denote the distribution of that random variable and the right-hand side is a convex combination of the two probability measures. From (1.3) we immediately get

(1.4) COROLLARY. *All stationary distributions have the form* $\theta\delta_\varnothing + (1-\theta)\xi_\infty^1$.

The result for $P_\lambda(\Omega_\infty) > 0$ follows from (1.3). When $P_\lambda(\Omega_\infty) = 0$, it follows from self-duality that $\xi_t^1 \Rightarrow \delta_\varnothing$ and from attractiveness that

$\xi_t^A \Rightarrow \delta_\varnothing$ for all $A \subset \mathbf{Z}^d$; so $\delta_\varnothing$ is the unique stationary distribution in this case.

(1.3) describes the limiting behavior of the finite-dimensional distribution. Our next result gives a more global description of ξ_t^0. Let

$$H_t = \bigcup_{t \geq 0} \xi_t^0 \quad \text{and} \quad K_t = \xi_t^0 \cup (\xi_t^1)^c .$$

H_t is the set of sites hit by time t. K_t is the coupled region. If we let $\xi_t(x) = 1$ when $x \in \xi_t$ and $\xi_t(x) = 0$ when $x \notin \xi_t$, then $\xi_t^0(x) = \xi_t^1(x)$ for $x \in K_t$. The next result shows that when ξ_t^0 does not die out, $t^{-1}(H_t \cap K_t)$ has an asymptotic shape. To state the result it is convenient to let $Q = [-\frac{1}{2}, \frac{1}{2}]^d$,

$$\overline{H}_t = \bigcup_{x \in H_t} (x + Q) \quad \text{and} \quad \overline{K}_t = \bigcup_{x \in K_t} (x + Q) .$$

(1.5) Shape Theorem. *Suppose $\lambda > \lambda_f$. There is a convex set D so that if $\omega \in \Omega_\infty$ and $\varepsilon > 0$, then for $t \geq t_0(\varepsilon, \omega)$*

$$(1 - \varepsilon)tD \subset (\overline{H}_t \cap \overline{K}_t), \qquad \overline{H}_t \subset (1 + \varepsilon)tD .$$

At this point several remarks are in order. First, we must intersect with K_t because $K_t \supset \{x : \xi_t^1(x) = 0\}$. Second, we cannot have coupling on a much larger set since $\xi_t^0(x) = 0$ on H_t^c. Third, when the processes are constructed on the graphical representation, $\xi_t^A = \bigcup_{x \in A} \xi_t^x$; so (1.5) generalizes easily to finite initial configurations A.

(1.5) is proved by checking the hypotheses of a general result of Durrett and Griffeath [**10**]. To do this we have to show that if $\tau = \inf\{t : \xi_t^0 = \varnothing\}$ then

$$(1.6) \qquad Ce^{-bt} \geq P(t < \tau < \infty) = P(x \in \xi_\infty^1) - P(x \in \xi_t^1),$$

the equality following from duality. The reader should note that (1.6) implies that $\xi_t^1 \Rightarrow \xi_\infty^1$ exponentially rapidly. With this established it is straightforward to imitate the proof in Section 13 of Durrett [**5**] to conclude

(1.7) Strong law of large numbers. *If $\lambda > \lambda_f$ then as $t \to \infty$,* $|\xi_t^0|/t^d \to \rho|G| 1_{\Omega_\infty}$ a.s.

Here $\rho = P(0 \in \xi_\infty^1)$ and $|G| =$ the volume of G. It would be interesting to prove a corresponding central limit theorem. See Galves and

Presutti [16] and Kuczek [27] for proofs of a central limit theorem for $r_t = \sup \xi_t^{(-\infty,0]}$ in $d = 1$.

The key to the proofs of (1.2), (1.3), and (1.5) is $(*)$. Bezuidenhout and Grimmett's proof of this result is described in Sections 3 and 4. To orient and motivate the reader, we begin by describing the general features of their construction in Section 2 and show how it implies (1.2). (1.3) and (1.5) are proved in Section 5.

2. The big picture. The first step in making $(*)$ precise is to define the percolation process. Let $\mathscr{L} = \{(m, n) \in \mathbf{Z}^2 : m+n \text{ is even}\}$, and let $\eta(z)$, $z \in \mathscr{L}$, be independent random variables with $P(\eta(z) = 1) = p$ and $P(\eta(z) = 0) = 1 - p$. When $\eta(z) = 1$ (0) we say that the site z is open (closed). We think of $\mathscr{L}$ as a graph with arcs connecting (m, n) to $(m+1, n+1)$ and from (m, n) to $(m-1, n+1)$. With this in mind we say that y can be reached from x and write $x \to y$ if there is a sequence of open sites $x_0 = 0$, $x_1, \dots, x_n = y$ so that $x_m - x_{m-1} \in \{(1, 1), (-1, 1)\}$ for $1 \le m \le n$. Let $C_0 = \{z : 0 \to z\}$ be the cluster containing 0, the origin in $\mathbf{Z}^2$, and let $\Lambda_\infty = \{|C_0| = \infty\}$ be the event that "percolation occurs." It is known (see Durrett [5], Section 10 or [6], p. 86) that

(2.1) LEMMA. *If $p > 80/81$ then $P(\Lambda_\infty) > 0$.*

Let $B_0 = [-2L, 2L]^d \times [0, T]$ and $B_{m,n} = (4Lm, 50Tn) + B_0$ for $(m, n) \in \mathscr{L}$. Let $\bar{\xi}_t^{4L}$ be the process in which births outside $(-4L, 4L)^d$ are not allowed. Let $I = (-J, J)^d$ and call a site $(m, n) \in \mathscr{L}$ wet if $\xi_t \supset x + (-J, J)^d$ for some $(x, t) \in B_{m,n}$. The heart of the proof of $(*)$ is

($\heartsuit$) PROPOSITION. *Suppose $P_\lambda(\Omega_\infty) > 0$. If $\varepsilon > 0$ then J, L, and T can be chosen so that if $(0, 0)$ is wet then with probability $> 1 - \varepsilon$, $(1, 1)$ and $(-1, 1)$ will also be wet.*

See Figure 1 for a picture. We use the process $\bar{\xi}_t^{4L}$ because the regions $(4Lm, 50Tn) + ((-4L, 4L)^d \times \mathbf{R})$, $m \in 2\mathbf{Z} - n$, are disjoint and hence what happens in the graphical representation of them will be independent. Using ($\heartsuit$) and induction it follows (a proof will be given at the end of Section 4) that the collection of wet sites dominates oriented percolation with probability $p = 1 - \varepsilon$. This leads easily to

(2.2) THEOREM. $P_{\lambda_f}(\Omega_\infty) = 0$.

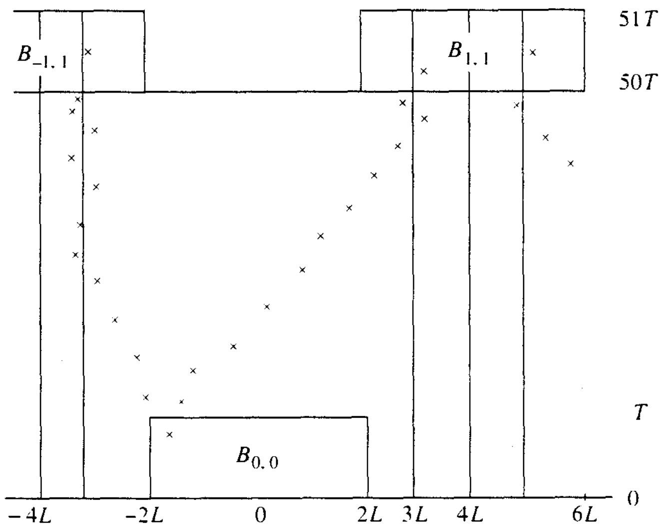


FIGURE 1

PROOF. Suppose $P_{\lambda_f}(\Omega_\infty) > 0$. Let $\varepsilon = 0.01$ in (♡) and pick J, L, and T so that the conclusion holds. Let $\lambda < \lambda_f$ and construct the graphical representation for rate λ by flipping coins with probability λ/λ_f of heads to determine which arrows to keep. Since the total number of arrows in $(-4L, 4L)^d \times [0, 51T]$ has a Poisson distribution, it is easy to see that if λ is close enough to λ_f the event in question has probability $> 80/81$. This and (2.1) show that the contact process with parameter λ dominates an oriented percolation process with $P(\Lambda_\infty) > 0$ and hence has $P_\lambda(\Omega_\infty) > 0$ contradicting the fact that $\lambda < \lambda_f$.

A second immediate consequence of the construction is

(2.3) THEOREM. $\lambda_f(\mathbf{Z}^d) = \lim_{M\to\infty} \lambda_f(\mathbf{Z} \times \{-M, \dots, M\}^{d-1})$.

PROOF. The sequence on the right-hand side is decreasing; so the limit exists and is easily seen to be $\geq \lambda_c(\mathbf{Z}^d)$. To prove the other inequality let $\lambda > \lambda_c(\mathbf{Z}^d)$, and observe that in the entire construction no births are allowed outside $\mathbf{Z} \times (-4L, 4L)^{d-1}$.

REMARK. The last result is false if λ_f is replaced by λ_e. The process in Durrett and Gray [9] has $\lambda_e(\mathbf{Z}^2) < \infty$ but $\lambda_e(\mathbf{Z} \times \{-M, \dots, M\}) = \infty$ for all M.

3. The construction, part I. In this section we will do the tricky part of the construction. The reader will see in the next section that once (♣) is established it is reasonably straightforward to get (♡). Throughout the section we will suppose $P_\lambda(\Omega_\infty) > 0$. Let $\gamma \in (0, 1)$ and $\delta = \gamma^2/3$.

(3.1) LEMMA. *If J is large enough and $I = (-J, J)^d$ then $P(\xi_t^I \neq \varnothing$ for all $t) \geq 1 - \delta$.*

PROOF. Let $\tau^0 = \inf\{t : \xi_t^0 = \varnothing\}$. Pick t so that $P(t < \tau^0 < \infty)/P(\tau^0 > t) \leq \delta$. The Markov property implies that

$$P(t < \tau^0 < \infty) = \mathrm{E}(P(\tau^{\xi_t} < \infty) ; \tau^0 > t).$$

Since $P(|\xi_t| < \infty) = 1$ it follows from the choice of t that $P(\tau^A < \infty) \leq \delta$ for some finite set A. If we pick J large enough, $(-J, J)^d \supset A$ and the result follows.

Let $\bar{\xi}_t^{I,K}$ be the process with $\bar{\xi}_0^{I,K} = I$ and in which births outside $H = [-K, K]^d$ are not allowed, and let

$$\mu_U = \int_0^U |\bar{\xi}_t^{I,K} \cap \partial H| \, dt$$

where ∂H is the boundary of H. Our aim will be to show

(♣) PROPOSITION. *If $M, N < \infty$, then K and U can be chosen so that $|\bar{\xi}_U^{I,K}| \geq M$ and $\mu_U \geq N$ each have probability $\geq 1 - \gamma$.*

To keep the logic in the proof straight, we will first pick M and N large enough so that when the events in (♣) occur, then with high probability (i) there is an $x \in \bar{\xi}_V^{I,K}$ so that $x + I \subset \bar{\xi}_{V+1}^{I,K+J}$ and (ii) we can find a fully occupied copy of I in $\bar{\xi}_t^{I,K+2J}$ in $(\overline{H} - H) \times [0, U]$ where $\overline{H} = [-K - 2J, K + 2J]^d$.

(3.2) LEMMA. *If M is large and $|\xi_0| \geq M$, then $P(\xi_1 \supset x + I$ for some $x \in \xi_0) \geq 1 - \gamma$.*

PROOF. We can find $\zeta \subset \xi_0$ with $|\zeta| \geq [M/(2J)^d]$, where $[x] =$ the greatest integer $\leq x$, so that $x + (-J, J)^d$, $x \in \zeta$, are disjoint. Let $\bar{\xi}_t^x$ be the process (constructed on the original graphical representation) with $\bar{\xi}_0^x = \{x\}$ and in which births ouside $x + (-J, J)^d$ are not allowed. $P(\bar{\xi}_1^x \supset x + I) = \beta > 0$ and success for different $x \in \zeta$ are independent.

(3.3) LEMMA. *There is a constant N so that if S is a subset of $\{0\} \times \mathbf{Z}^{d-1} \times [0, \infty)$ with $|S| \geq N/(2d)$ then with probability $\geq 1 - \gamma$*

there is an arrow from some $(x, t) \in S$ *to* $(x + e_1, t)$ *and the process starting with* $x + e_1$ *occupied at time* t *and no births outside* $I' = x + ((0, 2J) \times (-J, J)^{d-1}) = x + (J, 0, \ldots, 0) + I$ *contains* I' *at time* $t + 1$.

PROOF. Let $\mathbf{H} = [0, 1) \cup [2, 3) \cup \cdots$. Without loss of generality

$$S \subset \{0\} \times (2J)\mathbf{Z}^{d-1} \times \mathbf{H} \quad \text{and} \quad |S| \geq N' = \left(\frac{N/(2d)}{2(2J)}\right)^d .$$

Now if $A \subset [0, 1)$ the probability that a rate λ Poisson process has at least one arrival in A is

$$(1 - e^{-\lambda|A|}) = \int_0^{\lambda|A|} e^{-x}\,dx \geq \lambda|A|e^{-\lambda} \quad (\text{since } |A| \leq 1).$$

Let $X(z, k) = 1$ if there is a birth from some $(x, t) \in \{z\} \times [k, k+1)$ to $(x + e_1, t)$, 0 otherwise. The $X(z, k)$ are independent with $\mathrm{E}X(z, k) \geq \mathrm{var}(X(z, k))$; so if $W = \Sigma X(k, z)$ it follows from Chebyshev's inequality that $P(W < \mathrm{E}W/2) \leq 4/(\mathrm{E}W)$. Since $\mathrm{E}W \geq \lambda e^{-\lambda} N'$, we have shown that with high probability there are a lot of intervals in which there are arrows out of S. Because the intervals are separated in space and time, each one with an arrow gives us an i.i.d. chance of getting the cube we seek, and (3.3) follows.

Having chosen J, M, and N we turn now to the proof of (♣).

(3.4) LEMMA. *If* $|A| \leq M$ *then* $P(\tau^A < \infty) \geq (2d\lambda + 1)^{-M}$.

PROOF. The right-hand side gives the probability that all the particles die before giving birth.

Recall that $\delta = \gamma^2/3$ and pick V so that

$$P(V < \tau^I < \infty) \leq \delta(2d\lambda + 1)^{-M} e^{-\lambda N}.$$

This is more than we need for the proof of (3.5). The reader will see the reason for the choice of V in (3.8) and for the relationship between δ and γ in (3.9).

(3.5) LEMMA. $P(|\xi_V^I| \geq M) \geq 1 - 2\delta$.

PROOF. By (3.4) and the choice of V, we must have $P(0 < |\xi_V^I| \leq M) \leq \delta e^{-\lambda N} < \delta$. The result now follows from (3.1).

Our next step is to localize the construction in space.

(3.6) LEMMA. *If K is large then $P(\xi_t^I \subset (-K, K)^d$ for all $t \le V) \ge 1 - \delta$.*

PROOF. Let ζ_t be the contact process with no deaths. Comparison with a branching process shows that $E|\zeta_V| \le |I| \exp(2d\lambda V)$. Now use Chebyshev's inequality to bound $P(|\zeta_V| \ge K)$ and observe that ζ_V is connected.

PROOF OF (♣). It follows from (3.5) and (3.6) that

$$P(|\bar{\xi}_V^{I,K}| \ge M) \ge 1 - 3\delta = 1 - \gamma^2 > 1 - \gamma, \tag{3.7}$$

proving the first claim. Since $P(|\bar{\xi}_t^{I,K}| > 0$ for all $t) = 0$ and $t \to P(|\bar{\xi}_t^{I,K}| \ge M)$ is continuous, we can pick $U \ge V$ so that $P(|\bar{\xi}_U^{I,K}| \ge M) = 1 - \gamma$. When

$$\mu_U = \int_0^U |\bar{\xi}_t^{I,K} \cap \partial H| \, dt \le N,$$

the probability that no birth outside the box will be attempted is $\ge e^{-\lambda N}$. Combining this with (3.4) gives

(3.8)
$$P(V < \tau^I < \infty \mid |\bar{\xi}_V^{I,K}| > 0, |\bar{\xi}_U^{I,K}| \le M, \mu_U \le N) \ge (2d\lambda + 1)^{-M} e^{-\lambda N};$$

so it follows from the choice of V that

$$P(|\bar{\xi}_V^{I,K}| > 0, |\bar{\xi}_U^{I,K}| \le M, \mu_U \le N) \le \delta.$$

By (3.1) and (3.6), $P(|\bar{\xi}_V^{I,K}| = 0) \le 2\delta$. Harris' inequality (see note below), and the choice of U imply

$$\begin{aligned} \gamma^2 = 3\delta &\ge P(|\bar{\xi}_U^{I,K}| \le M, \mu_U \le N) \\ &\ge P(|\bar{\xi}_U^{I,K}| \le M) P(\mu_U \le N) = \gamma P(\mu_U \le N). \end{aligned} \tag{3.9}$$

This shows $P(\mu_U \le N) \le \gamma$ and proves the second claim.

(♪) Harris' inequality (see, e.g., Durrett [6, p. 129]) says that if $X_1, \dots, X_n$ are independent random variables and f and g are bounded nonincreasing functions defined on $\mathbf{R}^n$ then $E(f(X)g(X)) \ge Ef(X)Eg(X)$. [The result is usually stated for nondecreasing functions, but multiplying both functions by -1 gives the version we use here.] We want to apply this result when f and g are the indicator functions of $|\bar{\xi}_U^{I,K}| \le M$ and $\mu_U \le N$. To extend this result from a finite number of independent random variables to the graphical representation, let $\Gamma_\varepsilon = \{(M, n\varepsilon) : m \in \mathbf{Z}^d\}$ and consider Γ_ε as a graph with oriented bonds from $(x, n\varepsilon)$ to $(x, (n+1)\varepsilon)$ that are open with probability $1-\varepsilon$, and oriented bonds from $(x, n\varepsilon)$ to $(x+y, (n+1)\varepsilon)$ that are open with probability $\lambda\varepsilon$ when $|y|=1$. Letting $\varepsilon \to 0$ this random graph converges to the graphical representation. (See Durrett [**6**], Section 5c for more details.) Harris' inequality implies that the desired result holds for the approximating systems. Letting $\varepsilon \to 0$ and noticing that $P(\mu_U = N) = 0$ and $|\bar{\xi}_U^{I,K}|$ is integer-valued gives the result we used.

4. The construction, part II. (♣) together with (3.2) and (3.3) shows that given an occupied cube, we will with high probability get an occupied cube at the top of $H \times [0, U]$ and one on the side. The first step in bootstrapping this to (♡) is a generalization of the "square root trick" of percolation (see, e.g., Durrett [**6**, p. 131]).

(4.1) LEMMA. *Suppose A_i, $i \in I$, are increasing events that all have the same probability and let $A = \bigcup_i A_i$. Then $P(A_i) \geq 1 - (1 - P(A))^{1/|I|}$, where $|I|$ is the number of events.*

PROOF. Set theory, Harris' inequality, and the fact that $P(A_i) = P(A_1)$ imply

$$1 - P(A) = P\left(\bigcap_i A_i^c\right) \geq \prod_i P(A_i^c) = (1 - P(A_1))^{|I|}.$$

Rearranging gives the desired inequality.

Combining (4.1) and (♣) gives

(4.2) LEMMA. *Let $D = 2^d d$. If $M, N < \infty$, then K and U can be chosen so that*

$$|\bar{\xi}_U^{I,K} \cap \{x \geq 0\}| \geq \frac{M}{2d}$$

and

$$\frac{N}{D} \leq \bar{\mu}_U \equiv \int_0^U |\bar{\xi}_t^{I,K} \cap \partial H \cap \{x_1 = K, x_i \geq 0, \ i > 1\}| \, dt$$

each have probability $\geq 1 - \gamma^{1/D}$.

PROOF. For $\nu \subset \{1, \dots, d\}$ let $O_\nu = \{x_i \geq 0, i \in \nu, x_j \leq 0, j \notin \nu\}$, and $A_\nu = \{|\bar{\xi}_U^{I,K} \cap O_\nu| \geq M/2^d\}$, and observe that if $|\bar{\xi}_U^{I,K}| \geq M$, at least one of the 2^d events A_ν will occur. The second conclusion is similar but there are $D = 2^d d$ possibilities to consider. (For d

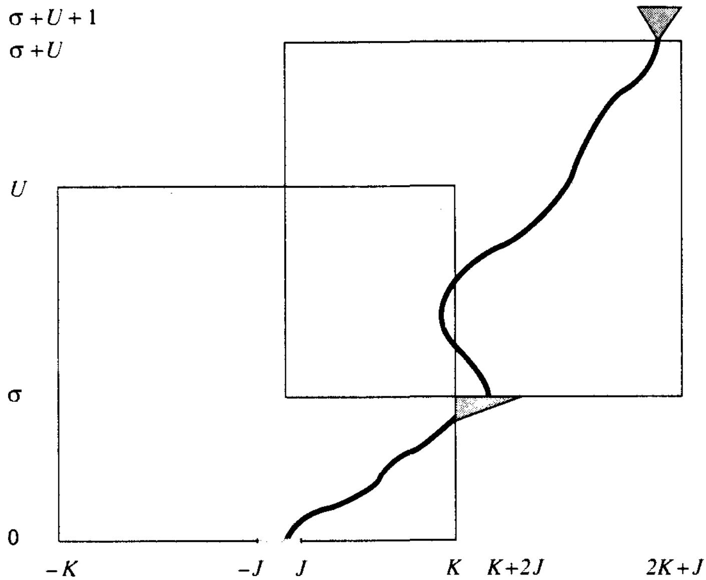


FIGURE 2

choices of j we have $x_j = \pm K$, with 2^{d-1} choices of signs for the other coordinates.)

(4.2) allows us to have the occupied copies of I on the side or top of $H \times [0, U]$ in any orthant we want and leads easily to

(4.3) LEMMA. *Let* $G = [K + J, 2K + J] \times [0, 2K]^{d-1} \times [U + 1, 2U + 2]$,

$$P(\bar{\xi}_t^{I, 2K+2J} \supset x + I \text{ for some } (x, t) \in G) \geq 1 - 4\gamma^{1/D}.$$

PROOF. See Figure 2 for a picture. Let $\sigma = \inf\{t : \bar{\xi}_t^{I, K+2J} \supset y + I$ for some y with $y_1 = K + J$ and $y_i \geq 0$ for $i > 1\}$. By (4.2) and (3.3),

$$P(\sigma > U + 1) \geq 1 - \gamma^{1/D} - \gamma \geq 1 - 2\gamma^{1/D} \qquad (\gamma < 1)$$

and the desired result follows by applying (4.2) and (3.2) to the process that starts with $y + I$ occupied at time σ and does not allow births outside $y + [-2K, 2K]^d$.

Using (4.3) we can drive our cube wherever we want to go. To prove (♡) we let $L = 8K$ and $T = 2U + 2$. We consider $(3L, 0, \ldots, 0)$ as our "target" and use it to choose the next direction to move. For example, if the center of the cube is at x with $x_1 < 3L$, $x_2 > 0$, $x_3 < 0$, we choose to put the next center $\bar{x} \in x + \{[K+J, 2K+J] \times [-2K, 0] \times [0, 2K]\}$. If $x_1 > 3L$ then we put the center in $x + [-2K - J, -K - J] \times \cdots$. Since $J \leq K$, each moves changes x_1 by an amount in $[K, 3K]$ and uses up between $T/2$ and T units of time. If we start with x_1 in $[-2L, 2L] = [-16K, 16K]$ (at some time $\leq T$) and want to get into $[2L, 4L] = [16K, 32K]$, at most 32 moves are required, and we can achieve $x_1 \geq 3L$ by time $41T$.

At the beginning of the construction the center of our cube has $|x_2| \leq 2L = 16K$. It is easy to see that if $x_2 > 0$ then $x_2 \geq \bar{x}_2 \geq -2K = -0.25L$; so we will always have $|x_i| \leq 2L$ for $i \geq 2$. For the first coordinate of the center we observe that it is nondecreasing until $3L$ is reached, and at that time $x_1 < 3L + 3K = 3.375L$. Repeating the argument for x_2 shows that we have $|x_1 - 3L| < 3K$ after x_1 first reaches $3L$. To see why we pick $L = 8K$ observe that when the center of the cube is at x, no births are allowed outside $x + [-2K - 2J, 2K + 2J]^d$. Now $J \leq K$, so if $|x_1 - 3L| \leq 3K$ and $|x_i| \leq 2L$ for $i > 1$, this is in $[2L, 4L] \times [-4L, 4L]^{d-1}$ with a little room to spare. Similar arguments show that before x_1 reaches $3L$ no births are allowed outside $[-2L - 4K, 3L + 4K] \times [-4L, 4L]^{d-1}$.

The arguments in the last paragraph show that if all goes well we can keep the center of the cube and the space time boxes needed to move it inside

$$R_{0,0} = ([-4L, 4L]^d \times [0, 50T]) \cup (\{[-4L, -2L] \cup [2L, 4L]\} \times [4L, 4L]^{d-1} \times [50T, 51T]).$$

(The reader will see the need for this funny-shaped region in the induction argument.) At most 100 moves are required; so if the γ in the last section is small enough then with probability close to one all will go well. To finish up now we will sketch the induction argument needed to prove

(♦) PROPOSITION. *Call a site $(m, n) \in \mathscr{L}$ wet if $\xi_t^I \supset x + (-J, J)^d$ for some $(x, t) \in B_{m,n}$. We can define independent random variables $\eta(z)$ with $P(\eta(z) = 1) = 1 - \varepsilon$ so that the wet sites contain C_0 for the oriented percolation generated by η.*

PROOF. We only have to define $\eta(m, n)$ when $|m| \leq n$. Suppose $n \geq 0$ and $\eta(j, k)$ have been defined for $k < n$. If $\{(m-1, n-1), (m+1, n-1)\} \cap C_0 = \varnothing$, flip a coin with a probability $1 - \varepsilon$ of heads to define $\tilde{\eta}(m, n)$. If at least one of the sites $(m-1, n-1)$, $(m+1, n-1)$ is in C_0, there will be at least one cube in $B_{m,n}$ that we can use for the starting point of our construction. If there are two, pick the one that occurs at the latter time. Let $\tilde{\eta}(m, n) = 1$ if the construction is successful starting with this cube. If we condition on everything that has happened in $R_{j,k} = (4Lj, 50Tk) + R_{0,0}$ with $k < n$ up to the appearance of the chosen cubes then the $\tilde{\eta}(m, n)$ are independent. (For this step we need $R_{0,0}$ instead of $[-4L, 4L]^d \times [0, 51T]$.) By introducing independent coin flips we can find i.i.d. $\eta \le \tilde{\eta}$ with $P(\eta = 1) = 1 - \varepsilon$ and the proof is complete.

5. Complete convergence and shape theorems. In this section we will prove (1.3) and (1.5). The first step is to identify what we need to show. The key to (1.3) is a result of Griffeath [**18**].

(5.1) LEMMA. *Let $\tilde{\xi}_t^B$ be an independent copy of the contact process. If*

(a) $P(\xi_t^A \cap \tilde{\xi}_t^B = \varnothing, \xi_t^A \neq \varnothing, \tilde{\xi}_t^B \neq \varnothing) \to 0$

then we get the conclusion of the complete convergence theorem:

$$P(\xi_{2t}^A \cap B \neq \varnothing) \to P(\tau^A = \infty)P(\xi_\infty^1 \cap B \neq \varnothing).$$

PROOF. $\{\xi_{2t}^A \cap B \neq \varnothing\} = \{\xi_t^A \cap \hat{\xi}_t^{(B,2t)} \neq \varnothing\}$ and ξ_t^A, $\hat{\xi}_t^{(B,2t)}$ are independent. Since we can only have $\xi_t^A \cap \hat{\xi}_t^{(B,2t)} \neq \varnothing$ when both sets are nonempty, the hypothesis implies

$$P(\xi_{2t}^A \cap B \neq \varnothing) - P(\tau^A > t)P(\tau^B > t) \to 0.$$

Since $P(\tau^B = \infty) = P(\xi_\infty^1 \cap B \neq \varnothing)$ the desired result follows.

REMARK. Since $A \to P(\xi_{2t}^A \cap B \neq \varnothing)$ is increasing and bounded above by $P(\xi_{2t}^1 \cap B \neq \varnothing)$, it suffices to prove the complete convergence theorem for finite sets A.

As advertised in the introduction we will prove (1.5) by checking the hypotheses of a result of Durrett and Griffeath [**10**].

(5.2) THEOREM. *Let $\tau = \tau^{\{0\}}$ and $t(x) = \inf\{x : x \in \xi_t^0\}$. Suppose there are constants a, b, $C \in (0, \infty)$ so that*

(b) $P(t < \tau < \infty) \le Ce^{-bt}$ *and*

(c) $P(t(x) > t, \tau < \infty) \le Ce^{-bt}$ *for* $|x| < at$.

Then there is a convex set G *so that if* $\omega \in \{\tau = \infty\}$ *then for all* $\varepsilon > 0$

$$(1-\varepsilon)tG \subset \overline{H}_t \subset (1+\varepsilon)tG \quad \text{for } t \geq t_0(\varepsilon, \omega).$$

If, in addition,

(d) $P(x \notin \overline{K}_t, \tau = \infty) \leq Ce^{-bt}$ *for* $|x| < at$,

then for $\omega \in \{\tau = \infty\}$ *and any* $\varepsilon > 0$

$$(1-\varepsilon)tG \subset \overline{H}_t \cap \overline{K}_t \subset (1+\varepsilon)tG \quad \text{for } t \geq t_1(\varepsilon, \omega).$$

REMARK. To explain why we are proving (1.3) and (1.5) in parallel note that (d) is implied by $P(\xi^0_{t/2} \cap \tilde{\xi}^x_{t/2} = \varnothing, \xi^0_{t/2} \neq \varnothing, \tilde{\xi}^x_{t/2} \neq \varnothing) \leq C\exp(-bt)$ for $|x| < at$.

To prove (a)–(d) we will use (♦) to get a lower bound on ξ^A_t when it survives. The approach is the same as the proofs in Section 5 of Durrett and Schonmann [**13**] and Section 5 of Bramson, Ding, and Durrett [**2**]; so we will concentrate on explaining the ideas and not bother to spell out all the details. To eliminate ...'s and to facilitate mental pictures we will suppose that $d = 2$. The reader will see that argument generalizes easily to $d > 2$.

The main ideas behind the proof are

(i) it is easy to prove what we want for oriented percolation with $p \geq 1 - \varepsilon_0$;

(ii) (♦) tells us that the contact process dominates oriented percolation with $p = 1 - \varepsilon_0$.

(iii) By repeatedly trying the "renormalized site construction" (or r.s.c. for short) from the proof of (♦), we will eventually get percolation or end up with $\xi^A_t = \varnothing$ and in either case we are happy.

Having announced the philosophy, we plunge into the details. Logically the first thing to do is to pick ε_0. Let C_0 be the cluster containing 0 in oriented site percolation with parameter p. Let $W_n = \{m : (m, n) \in C_0\}$, $\Lambda_\infty = \{|C_0| = \infty\} = \{W_n \neq \varnothing \text{ for all } n\}$, $\rho(p) = P(\Lambda_\infty)$, $\ell_n = \inf W_n$, and $r_n = \sup W_n$. Well-known results for oriented percolation (see Durrett [**5**, Sections 3 and 13], or [**6**, Chapters 4 and 11]) show that

(5.3) a.s. on Λ_∞ we have

$$r_n/n \to \alpha(p),\ \ell_n/n \to -\alpha(p), \text{ and } |W_n|/n \to \alpha(p)\rho(p).$$

(5.4) If $p > 1 - \varepsilon_0$, then $\alpha(p)$, $\rho(p) > 0.99$.

The reader should note that on Λ_∞, the length of the interval $[\ell_n, r_n]$ is $\sim 2\alpha(p)n$ but only every other point is in $\mathcal{L}$, so W_n fills up most of the interval.

Let $\varepsilon = \varepsilon_0/2$ and pick j, L, and T so that (♡) from Section 2 holds. The next step is to look for an occupied copy of $I = (-J, J)^d$ to try the r.s.c. For $z \in \mathbf{Z}^d$, let $I_z = 4Lz + I$. There is a $\delta > 0$ so that $P(\xi_1^x \supset I_z$ for some $z) \geq \delta$; so after at most a geometric number of trials κ we will have $\xi_\kappa^A = \varnothing$ or we will have found an $I_z \subset \xi_{\kappa+1}^A$. In the second case we try the r.s.c. in $z_1 + (\mathbf{R} \times [-4L, 4L])$. If we get percolation we are happy. If not, we wait until $W_n = \varnothing$ and we look again for an occupied I_z to try again. After a geometric number of repetitions we will have $\xi_s^A = \varnothing$ or a successful r.s.c. starting from some I_ζ. It follows from estimates on pages 1031–1032 in Durrett [5] that if σ_0 is the amount of time used up in this part of the construction then

$$(5.5) \qquad P(\sigma_0 > t) \leq Ce^{-bt}.$$

Since at time σ_0 we either have $\xi_{\sigma_0}^A = \varnothing$ or know that $\tau^A = \infty$ or know that $\tau^A = \infty$, it follows that

$$(5.6) \quad P(t < \tau^A < \infty) \leq Ce^{-bt} \qquad \text{(with constants independent of } A\text{)}$$

proving (b). See Section 12 of Durrett [5] for more details. For later purposes we need to know that ζ is not too far from 0. By comparing with a contact process in which there are no deaths and using (5.5) it is easy to show that if $a = L/(1000T)$ then

$$(5.7) \qquad P(|\zeta| > at) \leq Ce^{-bt}.$$

The proof so far is valid in any dimension. For the next step we restrict our attention to $d = 2$. Let $\nu_1 = [\![t/((d+1)50T)]\!]$ where $[\![x]\!]$ is the largest even integer $\leq x$. (We pick an even time because then the wet sites are $\subset 2\mathbf{Z}$.) We run the percolation process for ν_1 units of time and then we use the wet sites to grow in the second direction, i.e., starting with each wet site m we try the r.s.c. in $(\zeta_1 + 4mL) + ((-4L, 4L) \times \mathbf{R})$. ($\zeta$ is the center of the successful cube.) With high probability (i.e., $\geq 1 - C\exp(-bt)$) the number of sources is $> (0.99)^3\nu$ (by (5.4) we expect to have $(0.99)^2$), and we will get $> (0.99)^5\nu$ successful r.s.c.'s. Let $\sigma_1 = 50T\sigma_1$. We run the new oriented percolations for

$$\nu_2 = [\![(t - \sigma_0 - \sigma_1)/(50T)]\!] - 2$$

time steps. (5.5) implies that $\nu_2 \geq \nu_1$ with high probability.

Let $V \equiv [-\nu_1/2, \nu_1/2] \times [-\nu_2/2, \nu_2/2]$ and let W be the set of $v \in \zeta + V$ that are wet in the percolation construction, i.e., contain an occupied copy of I. The arguments above show that $|W| \geq (0.99)^7 \nu_1 \nu_2 \geq 0.93\nu_1\nu_2$. With the last result in hand it is easy to prove (a), (c), and (d). To prove (a) we observe that if ξ_t^A and $\tilde{\xi}_t^B$ do not die out and W^A and W^B are the corresponding wet regions, then

(5.8) $$P(|W^A \cap W^B| \geq ct^2 \mid \xi_t^A \neq \varnothing, \tilde{\xi}_t^B \neq \varnothing) \leq Ce^{-bt}.$$

If we let $\sigma_2 = 50T\nu_2$, then the definition of ν_2 guarantees $t - (\sigma_0 + \sigma_1 + \sigma_2) \in [100T, 200T]$. Dividing space up into disjoint blocks and finding independent events we see that

(5.9) $$P(\xi_t^A \cap \tilde{\xi}_t^B = \varnothing \mid |W^A \cap \widetilde{W}^B| \geq ct^2) \leq \exp(-bt^2).$$

For more details in a more complicated situation where the parity of the intersection is important see Section 5 of Bramson, Ding, and Durrett [2].

(5.8) and (5.9) give us (a). To prove (d) we observe that the argument for (5.8) works when $A = \{0\}$, $B = \{x\}$, $|x| < 2at$, and $a = L/(1000T)$. To prove (c) we modify the first step of the construction for $\tilde{\xi}_t^x$ so that every time we look for our copy of I in some $\tilde{\xi}_t^{(x,s)}$, i.e., the process starting from x occupied at some time $s \geq 0$. In this way we will get survival of some $\tilde{\xi}_t^{(x,s)}$ and the proof given above guarantees that if $|x| < 2at$ then with high probability we will hit x by time $2t$.

References

1. C. Bezuidenhout and G. Grimmett, *The critical contact process dies out*, Ann. Probab., to appear.

2. M. Bramson, W. D. Ding, and R. Durrett, *Annihilating branching processes*, Stoch. Process. Appl., to appear.

3. R. C. Brower, M. A. Furman, and M. Moshe, *Critical exponents for the Reggeon quantum spin model*, Phys. Lett. **76B** (1978), 2113–2119.

4. R. Durrett, *On the growth of one-dimensional contact processes*, Ann. Probab. 8 (1980), 890–907.

5. R. Durrett, *Oriented percolation in two dimensions*, Ann. Probab. **12** (1984), 999–1040.

6. R. Durrett, *Lecture Notes on Particle Systems and Percolation*, Wadsworth, Pacific Grove, CA, 1988.

7. R. Durrett, *A new method for proving the existence of phase transitions*, preprint, 1989.

8. R. Durrett and L. Gray, *Some peculiar properties of a particle system with sexual reproduction*, In Tautu (1986).

9. R. Durrett and L. Gray, *Some peculiar properties of a particle system with sexual reproduction*, unpublished manuscript.

10. R. Durrett and D. Griffeath, *Contact processes in several dimensions*, Z. Für Wahr. **59** (1982), 535–552.

11. R. Durrett and D. Griffeath, *Supercritical contact processes on* $\mathbf{Z}$, Ann. Probab. **11** (1983), 1–15.

12. R. Durrett and Xiu-fang Liu, *The contact process on a finite set*, Ann. Probab. **16** (1988), 1158–1173.

13. R. Durrett and R. H. Schonmann, *Stochastic growth models*, In Kesten (1987).

14. R. Durrett and R. H. Schonmann, *The contact process on a finite set*, II, Ann. Probab. **16** (1988), 1570–1583.

15. R. Durrett, R. H. Schonmann, and N. Tanaka, *The contact process on a finite set, III: the critical case*, Ann. Probab. **17** (1989), 1303–1321.

16. A. Galves and E. Presutti, *Edge fluctuations for one dimensional contact processes*, Ann. Probab. **15**, 1131–1145.

17. P. Grassberger and A. de la Torre, *Reggeon field theory (Schlögls second model) on a lattice: Monte Carlo calculations of critical behavior*, Ann. Physics **122** (1979), 373–396.

18. D. Griffeath, *Limit theorems for nonergodic set-valued Markov processes*, Ann. Probab. **6** (1978), 379–387.

19. D. Griffeath, *The basic contact process*, Stochastic Process. Appl. **11** (1981), 151–185.

20. D. Griffeath, *The binary contact path process*, Ann. Probab. **11** (1983), 692–705.

21. T. E. Harris, *A lower bound for the critical probability in a certain percolation process*, Proc. Camb. Phil. Soc. **56** (1960), 13–20.

22. T. E. Harris, *Contact interactions on a lattice*, Ann. Probab. **2** (1974), 969–988.

23. T. E. Harris, *Additive set-valued Markov processes and graphical methods*, Ann. Probab. **6** (1978), 355–378.

24. R. Holley and T. Liggett, *The survival of contact processes*, Ann. Probab. **6** (1978), 198–206.

25. R. Holley and T. Liggett, *Generalized potlatch and smoothing processes*, Z. für Wahr. **55** (1981), 165–195.

26. H. Kesten (ed.), *Percolation Theory and the Ergodic theory of Interacting Particle Systems*, Vol. 8, IMA Volumes in Math. Appl., Springer-Verlag, New York, 1987.

27. T. Kuczek, *The central limit theorem for the right edge of supercritical oriented percolation*, Ann. Probab. **17** (1989), 1322–1332.

28. T. M. Liggett, *Interacting Particle Systems*, Springer-Verlag, New York, 1985.

29. P. Tautu (ed.), *Stochastic Spatial Processes*, Lecture Notes in Math. **1212**, Springer-Verlag, New York, 1986.

CORNELL UNIVERSITY
E-mail address: RTD@CornellA.bitnet

Lectures in Applied Mathematics
Volume 27 (1991)

Is the Contact Process Dead?

LAWRENCE F. GRAY

Abstract. The basic one-dimensional nearest-neighbor contact process has been the subject of many fruitful investigations during the past one and a half decades. Recent work of Bezuidenhout and Grimmett has solved what seems to be the last significant open question, and many researchers are turning their attention to other processes. But are we really done yet? In this report, we raise several simple questions about the contact process which are so fundamental that it is somewhat embarrassing that they have not been previously answered. It turns out that these questions are related: they all can be answered by using certain properties of *extremal paths* in the graphical representation of the contact process. Some of these properties are already known, but there are also several useful monotonicity properties that are announced here for the first time.

1. Introduction. The basic one-dimensional nearest-neighbor contact process could be called the simplest nontrivial interacting particle system, at least among models with continuous time. Unlocking its secrets has been a favorite pastime of workers in the field since it was first introduced by Harris [5]. Most of the attention has been focused on the equilibrium behavior. Many interesting results have been proved, culminating in the recent result of Bezuidenhout and Grimmett [1], which nails down the equilibrium behavior at the critical birth rate. A nice exposition of the current state of knowledge is to be found in Durrett's article [2], entitled "The Contact Process, 1974–1989" (is this an obituary notice?). To summarize, it is now known that the contact process's long-term behavior is very like that of a branching process: If the birth

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60K35.

Research for this preliminary report was supported in part by NSF Grant DMS 83-01080. The final version of this paper will be submitted for publication elsewhere.

rate parameter λ is strictly larger than some critical value λ_c, then starting from a single particle, there is a positive probability that the number of particles will grow forever; for all other values of λ (including $\lambda = \lambda_c$), the number of particles will reach 0 almost surely. The main difference between the contact process and branching processes seems to be that there is a formula for the critical birth rate for branching processes, whereas we only have bounds for the critical value of the contact process. There appears to be little hope of obtaining a precise numerical value for λ_c, so that is essentially the end of the story.

Or is it? In this article, I will discuss certain simple questions about the contact process in the hopes of showing that at age 15, the contact process still has some life in it. We will introduce these questions informally here. More precise statements, along with answers, will follow in subsequent sections.

The first question concerns what I call the "population profile." Assume that initially there is a single particle at the origin, and fix some time $t > 0$. The population profile is the function which tells, for each site x, the probability $\pi_t(x)$ that x is occupied at time t. This non-negative function is symmetric about the origin and goes to 0 as $|x|$ goes to ∞. Where is its maximum?

The second question concerns conditional probabilities. For each x and t, let $\mathscr{E}(x, t)$ be the event that x is occupied at time t. How does $P(\mathscr{E}(x, t)|\mathscr{E}(y, t))$ compare with $P(\mathscr{E}(x, t)|\mathscr{E}(y, t) \cap \mathscr{E}(z, t))$? The answer to this question is not as obvious as it seems at first.

The third and final question concerns a conjecture of Liggett [9]. Let $\sigma(A)$ be the probability that the number of particles never reaches 0 if A is the set of sites that are occupied initially. How does $\sigma(A \cup \{x\})$ depend on x? The answer to this question has implications for certain comparisons between the contact process and other models.

It seems to me that these questions are quite basic. It is not hard to imagine them arising naturally in applications. It turns out that there is a common thread: each of these questions is closely related to the properties of "extremal paths." In the next section, we will define paths and discuss some of their properties. In the remaining three sections, we will discuss applications of these properties to the three questions raised above.

2. Paths of the contact process. The most useful construction of the contact process is the graphical representation, due to Harris [7]. As shown in Figure 1, this consists of a random graph containing vertical half-lines directed upward, horizontal segments of unit length directed either to the right or to the left, and x's. Each vertical half-line is a time line associated with one of the sites in $\mathbf{Z}$. Shown are the time lines associated with $-3, -2, -1, 0, 1, 2, 3$. We call the horizontal segments *birth arrows*, or simply *arrows*. The points where the tails of the arrows intersect the time lines are called *birth points*. There are two types of birth points, depending on the direction of the corresponding birth arrows. For each direction, the corresponding birth points on a given time line form a Poisson point process with intensity equal to λ times Lebesgue measure, where λ is a positive parameter called the *birth rate*. Or in other words, the lengths of the intervals between successive birth points of one type on any one time line are independent exponentially distributed random variables with expected value $1/\lambda$. Similarly, the lengths of the intervals between points marked with an "x" on any given time line are independent exponentially distributed random variables with expected value 1. The points marked with x's are called *death points*. The entire collection of exponentially distributed random variables used to determine the locations of arrows and x's is independent. The probability space underlying the random graph will be denoted by $(\Omega, \mathcal{F}, P)$. All random quantities to be defined throughout the rest of this article will be defined on $(\Omega, \mathcal{F}, P)$.

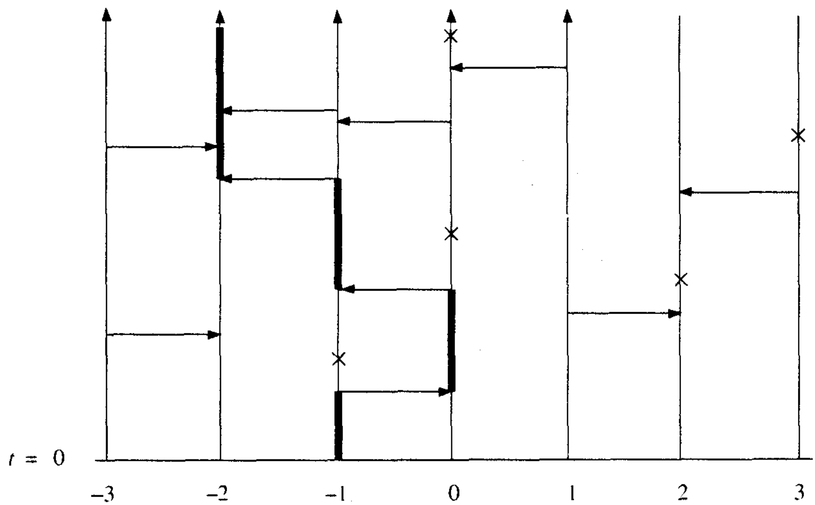


FIGURE 1

Before we use the random graph to define the contact process, we need to define *paths*. An $[s, t]$*-path* is a right continuous, piecewise constant function $\pi: [s, t] \to \mathbf{Z}$ which satisfies the following two conditions for all $u \in [s, t]$:

(i) $(\pi(u), u)$ is not a death point;
(ii) if $\pi(u^-) \neq \pi(u)$, then there is a birth point at $(\pi(u^-), u)$, and the head of the corresponding arrow points to $(\pi(u), u)$.

The image of one path is shown as a set of thickened lines in the picture. Note that this path, like all paths, travels upward along the vertical time lines, making jumps only at birth arrows (condition (ii)), and not passing over any death points (condition (i)). If π is an $[s, t]$-path such that $\pi(s) \in A$ and $\pi(t) \in B$, where A and B are subsets of $\mathbf{Z}$, then we say that π is a path *from* A *to* B.

The basic one-dimensional nearest-neighbor contact process is a Markov process whose state space is the collection of all subsets of $\mathbf{Z}$. Actually, there is one Markov process for each initial state. If the initial state is $A \subseteq \mathbf{Z}$, then we denote the corresponding process by (ξ_t^A). Given the graphical representation described above, the state of the contact process at time t is determined according to the following simple rule:

$$x \in \xi_t^A \quad \text{iff there is a } [0, t]\text{-path from } A \text{ to } \{x\}.$$

The contact process may be thought of as a model for the growth of a population. The points in $\mathbf{Z}$ represent *sites* that could be either *occupied* or *vacant*. The state of the system is the set of occupied sites. Each path represents a kind of a genealogical line. Death points are places (in space-time!) where these genealogical lines come to an end. Birth arrows represent places where the population may spread from one site to another, causing a branching of genealogical lines.

One important feature of the graphical construction of the contact process is that the entire family of processes $\{(\xi_t^A), A \subseteq \mathbf{Z}\}$ is defined on the same probability space $(\Omega, \mathcal{F}, P)$. This fact is useful in studying the behavior of the contact process with different initial states. Many other processes also have graphical representations. For a general treatment, including a proof that a well-defined process results from such a construction, see Gray [3] or Griffeath [4]. Another description of this graphical approach to the contact process is to be found in Durrett's article in this volume.

The process we have constructed is *one-dimensional* because the set of sites is the one-dimensional integer lattice. It is *nearest-neighbor* because the birth arrows all have unit length. These two properties imply the existence of special paths which we call *leftmost* and *rightmost* paths. Collectively, leftmost and rightmost paths will be called *extremal* paths. To define extremal paths, we need to put a partial ordering on functions from $[s, t]$ to $\mathbf{Z}$:

$$\pi_1 \le \pi_2 \quad \text{iff } \pi_1(u) \le \pi_2(u) \text{ for all } u \in [s, t].$$

It is easy to check from the graphical construction that whenever there exists an $[s, t]$-path from A to B, there is a leftmost $[s, t]$-path π_l and a rightmost path π_r, both from A to B, such that if π is any $[s, t]$-path from A to B, then $\pi_l \le \pi \le \pi_r$. Note the importance of the one-dimensional, nearest-neighbor character of the process. There are two $[0, t]$-paths from $\{-1\}$ to $\{-2\}$ in the picture (assuming that the top of the picture corresponds to time t). The path indicated is the leftmost such path.

Extremal paths have two properties which will be important to us. The first is a kind of space-time strong Markov property, which says that the presence of a rightmost path gives no information about what lies to the left of that path in the graphical representation, and similarly for leftmost paths. I do not really know who first discovered this property. It has been used in percolation theory for some time. A (somewhat tedious) proof in the context of interacting particle systems can be found in Gray [3]. Here is some necessary notation. For each right continuous, piecewise constant function from $[s, t]$ to $\mathbf{Z}$, let $\mathscr{F}_l(\pi)$ be the σ-field generated by the locations of all birth arrows and x's that lie strictly to the left of π in $\mathbf{Z} \times [s, t]$, and similarly let $\mathscr{F}_r(\pi)$ be the σ-field generated by the locations of all birth arrows and x's that lie strictly to the right of π.

PROPOSITION 0. *Let $\mathscr{E}$ be the event that there exists an $[s, t]$-path from A to B. If $\mathscr{E}$ occurs, let π_l and π_r be respectively the leftmost and rightmost $[s, t]$-paths from A to B. Let $\pi: [s, t] \to \mathbf{Z}$ be right continuous. Then for all events $\mathscr{A}_l \in \mathscr{F}_l(\pi)$ and $\mathscr{A}_r \in \mathscr{F}_r(\pi)$,*

$$P(\mathscr{A}_l | \mathscr{E}, \pi_r \ge \pi) = P(\mathscr{A}_l),$$

and

$$P(\mathscr{A}_r | \mathscr{E}, \pi_l \le \pi) = P(\mathscr{A}_r).$$

It is a consequence of the properties of Poisson point locations that the σ-fields $\mathscr{F}_l(\pi)$ and $\mathscr{F}_r(\pi)$ are independent. Thus, the preceding proposition is equivalent to the following statement:

$\mathscr{E}\cap\{\pi_l \le \pi\}$ is $\mathscr{F}_l(\pi)$-measurable and $\mathscr{E}\cap\{\pi_r \ge \pi\}$ is $\mathscr{F}_r(\pi)$-measurable.

The second property of extremal paths is new. It concerns a certain type of monotonicity that is quite useful. We first need a definition. Let $\Pi(s, t)$ be the collection of all piecewise constant right continuous functions from $[s, t]$ to $\mathbf{Z}$ (these are the kinds of functions that can be $[s, t]$-paths). A real-valued function φ defined on $\Pi(s, t)$ is called *increasing* if $\varphi(\pi_1) \le \varphi(\pi_2)$ for all $\pi_1, \pi_2 \in \Pi(s, t)$ such that $\pi_1 \le \pi_2$. If μ and ν are two distributions on $\Pi(s, t)$ (defined on an appropriate σ-field), we will say that μ is *stochastically to the left* of ν if for all increasing functions $\varphi: \Pi(s, t) \to \mathbf{Z}$,

$$\int \varphi(\pi)\, d\mu(\pi) \le \int \varphi(\pi)\, d\nu(\pi).$$

THEOREM 1. *Let A and B be two nonempty subsets of $\mathbf{Z}$. Choose x to be an integer that is strictly larger than $\sup A$. Let $\mathscr{D}$ be the event that an $[s, t]$-path exists from A to B, and let μ be the conditional distribution of the leftmost (rightmost) such path given $\mathscr{D}$. Also, let $\mathscr{E}$ be the event that an $[s, t]$-path exists from $A \cup \{x\}$ to B, and let ν be the conditional distribution of the leftmost (rightmost) such path given $\mathscr{E}$. Then μ is stochastically to the left of ν.*

There are several variations of this theorem. For example, the site x may be added to the left of A instead of to the right, or it may be added either to the left or the right of B. In all cases, increasing the size of one of the two sets A, B in a certain direction moves the conditional distributions of the extremal paths in that same direction.

There is also a variation of the theorem which concerns the effect of a "forbidden zone" on extremal paths. A forbidden zone is a region in the space-time graph that paths are not allowed to touch. (Of course, we are only interested in regions that are nice enough to avoid technical problems, but we will not be more specific here.)

THEOREM 2. *Let $\mathscr{D}$ be as in the preceding theorem, and let $\mathscr{D}_R$ be the event that there exists an $[s, t]$-path from A to B which lies strictly to the left of a given space-time region R. Assume that $P(\mathscr{D}_R) > 0$. Let π be the rightmost (or leftmost) $[s, t]$-path from A to B when $\mathscr{D}$ occurs. Then the conditional distribution of π given $\mathscr{D}_R$ is stochastically to the left of the conditional distribution of π given $\mathscr{D}$.*

Both of these theorems are proved by using the discrete time approximation to the contact process and a somewhat tricky induction. Details will appear in a future research article. In the remaining sections, we will show how these theorems can be applied to the questions raised in the introduction.

3. Population profiles. The *population profile* at time t is the function p_t defined as follows:

$$p_t(x) = P(x \in \xi_t^{\{0\}}).$$

It seems reasonable that this function should take its maximum at $x = 0$ and, in fact, be decreasing in $|x|$. We will show here how to use Theorem 1 to prove

THEOREM 3. *For all $t \geq 0$, the population profile p_t is a nonnegative, symmetric function of x which decreases in $|x|$.*

PROOF. Nonnegativity and symmetry are trivial properties of the population profile. Let

$$T = \inf\{t : p_t(y) > p_t(y+1) \text{ for some } y \geq 0\}.$$

Continuity considerations and the form of p_0 imply that $T > 0$. We wish to show that $T = \infty$. Suppose not. Then by continuity, $p_T(x)$ is decreasing in $|x|$ and there is a smallest nonnegative integer y such that

$$p_T(y) = p_T(y+1) \tag{1}$$

and

$$\dot{p}_T(y) \leq \dot{p}_T(y+1), \tag{2}$$

where $\dot{p}_T(x)$ is the derivative with respect to T of $p_T(x)$. We compute

$$\begin{aligned} \dot{p}_T(y) &= -p_T(y) + \lambda[P(y \notin \xi_T^{\{0\}}, y-1 \in \xi_T^{\{0\}}) \\ &\qquad + P(y \notin \xi_T^{\{0\}}, y+1 \in \xi_T^{\{0\}})] \\ &= -p_T(y+1) + \lambda[P(y \notin \xi_T^{\{0\}} | y-1 \in \xi_T^{\{0\}}) p_T(y-1) \\ &\qquad + P(y+1 \notin \xi_T^{\{0\}}, y \in \xi_T^{\{0\}})] \\ &\geq -p_T(y+1) + \lambda[P(y \notin \xi_T^{\{0\}} | y-1 \in \xi_T^{\{0\}}) p_T(y+2) \\ &\qquad + P(y+1 \notin \xi_T^{\{0\}}, y \in \xi_T^{\{0\}})]. \end{aligned} \tag{3}$$

So far, we have only used a standard formula for $\dot{p}_T$, the fact that $p_T(x)$ is decreasing in $|x|$, and (1). Suppose that we could prove that

$$P(y \notin \xi_T^{\{0\}} | y-1 \in \xi_T^{\{0\}}) > P(y+1 \notin \xi_T^{\{0\}} | y+2 \in \xi_T^{\{0\}}). \tag{4}$$

Then substituting the right side of (4) into (3) gives the standard expression for $\dot{p}_T(y+1)$. It then follows that $\dot{p}_T(y) > \dot{p}_T(y+1)$, which contradicts (2).

It remains to prove (4). Theorem 1 is precisely suited to proving such an inequality. We will also need Proposition 0. The argument that we give here is a little informal in some places, but there are standard ways of making it precise.

By symmetry and translation invariance, the left side of (4) equals

$$P(y+1 \notin \xi_T^{\{2y+1\}} \mid y+2 \in \xi_T^{\{2y+1\}}).$$

Let $\mathscr{D} = \{y+2 \in \xi_T^{\{0\}}\}$ and $\mathscr{E} = \{y+2 \in \xi_T^{\{2y+1\}}\}$. When $\mathscr{D}$ occurs, let π_1 be the rightmost $[0, t]$-path from 0 to $y+2$; and when $\mathscr{E}$ occurs, let π_2 be the rightmost $[0, t]$-path from $2y+1$ to $y+2$. Since $2y+1 > 0$, it follows from successive applications of Theorem 1 and one of its variations (the one in which the site x appears to the left of A) that the conditional distribution of π_1 given $\mathscr{D}$ is stochastically to the left of the conditional distribution of π_2 given $\mathscr{E}$, in the sense stated in the theorem. Note that if $\mathscr{D}$ occurs, the event $\{y+1 \notin \xi_T^{\{0\}}\}$ will happen if and only if there is no path from some part of π_1 to the space-time point $(y+1, T)$. A similar statement holds for the relationship between the event $\{y+1 \notin \xi_T^{\{2y+1\}}\}$ and the path π_2. But it is "stochastically farther" from $(y+1, T)$ to π_2 than it is to π_1 (see Figure 2). Proposition 0 says, in effect, that the presence of either of these paths has no effect on the distribution of birth arrows and death points in the region between the path and $(y+1, T)$. It follows that it is more likely that a path reaches from $(y+1, T)$ to π_1, conditioned on $\mathscr{D}$, than it is for a path to reach from $(y+1, T)$ to π_2, conditioned on $\mathscr{E}$. With a little more work, it can be shown that this inequality of probabilities is strict. The inequality in (4) now follows. □

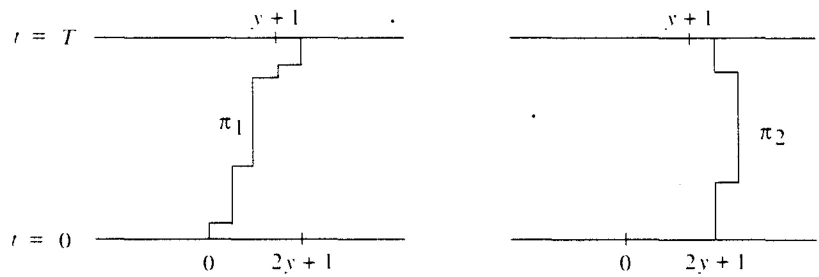


FIGURE 2

The preceding argument illustrates the way in which the properties of extremal paths can be used to prove certain inequalities involving conditional probabilities: first, the events that we are conditioning on are stated in terms of the presence of certain paths; then, an inequality is proved concerning the conditional distributions of the corresponding extremal paths, using a result like Theorem 1 or 2; finally it is shown, often by using Proposition 0, that such an inequality implies the desired inequality between the two conditional probabilities.

4. Conditional occupation. We will follow the scheme outlined at the end of the previous section to prove the following somewhat surprising inequality.

THEOREM 4. *For integers x and times t, let*

$$\mathscr{E}(x, t) = \{x \in \xi_T^{\{0\}}\}.$$

If $x < y < z$, then

$$P(\mathscr{E}(x, t)|\mathscr{E}(y, t)) \geq P(\mathscr{E}(x, t)|\mathscr{E}(y, t) \cap \mathscr{E}(z, t)) \tag{5}$$

$$\geq P(\mathscr{E}(x, t)). \tag{6}$$

The inequalilty in (6) is well known. It follows from the fact that the events $\mathscr{E}(x, t)$, $\mathscr{E}(y, t)$, $\mathscr{E}(z, t)$ and intersections thereof are positively correlated (see Harris [6] or Liggett's book [8]). We included it in order to emphasize the paradoxical appearance of (5).

We will outline the proof of the following, which is equivalent to (5):

$$P(\mathscr{E}(x, t)|\mathscr{E}(y, t)) \leq P(\mathscr{E}(x, t)|\mathscr{E}(y, t) \cap \mathscr{E}(z, t)^c). \tag{7}$$

The first step of the argument is to introduce the appropriate path: let π_y be the rightmost $[0, t]$-path from $\mathbf{Z}$ to $\{y\}$, which must exist if $\mathscr{E}(y, t)$ occurs. The second step is to apply one of the two theorems in Section 2. The appropriate choice here is Theorem 2. The idea is that the event $\mathscr{E}(z, t)^c$ is equivalent to the existence of a random forbidden zone R. If $\mathscr{E}(y, t) \cap \mathscr{E}(z, t)^c$ occurs, the path π_y must pass strictly to the left of R. It can be shown (this is similar to Proposition 0) that the presence of the forbidden zone R does not affect the locations of birth arrows and death points outside of R. This fact can be used to allow the application of Theorem 2, even though R is random: the conditional distribution of π_y given $\mathscr{E}(y, t) \cap \mathscr{E}(z, t)^c$ is stochastically to the left of the conditional distribution of π_y given $\mathscr{E}(y, t)$. The third step of the argument is identical to the last part of the proof of Theorem 3: the comparison between the two conditional distributions of π_y translates into (7) after an appropriate application of Proposition 0.

5. Liggett's conjecture. Let I be an interval of n consecutive sites:

$$I = \{x, x+1, \dots, x+n-1\}.$$

Let A be a finite subset of $\mathbf{Z}$ which contains the sites $x-1$ and $x+n$ but none of the sites in I. Thus, A is a finite set of sites with a "gap" I. Let

$$\sigma_t(y) = P(\xi_t^{A\cup\{y\}} \neq \varnothing).$$

This is the "survival probability" at time t when the initial state is $A \cup \{y\}$. We are interested in how $\sigma_t(y)$ depends on $y \in I$.

For simplicity, suppose that n is odd, and let $m = x + (n-1)/2$ be the middle site in I. Then one version of Liggett's conjecture (see Liggett [**9**]) is

THEOREM 5. *For* $k = 1, 2, \dots, (n-1)/2$,

$$\sigma_t(m) \geq (\sigma_t(m-k) + \sigma_t(m+k))/2. \tag{8}$$

The theorem says that after symmetrizing about the center site m, occupation of the center site increases the survival probability at least as much as occupation of any other site in the gap. Intuitively, m is the best site to occupy since it is the place at which "crowding" in the population is reduced to a minimum. Reducing crowding increases the survival probability, because it opens up more sites at which new particles can appear. This conjecture implies another conjecture which says that, in a certain sense, the basic one-dimensional nearest-neighbor contact process has the lowest "survivability" of any model in a class of related population models (the so-called nearest particle system). (See Liggett [**9**] for details.)

We will indicate briefly here how the theorems of Section 2 can be used to prove Theorem 5. Let us recast (8) in a form that is more amenable to our approach. Let

$$\mathscr{A} = \{\xi_t^A = \varnothing\}$$

and let

$$\sigma_t'(y) = P(\xi_t^y \neq \varnothing | \mathscr{A}).$$

Since $\xi_t^{A\cup\{y\}} = \xi_t^A \cup \xi_t^y$ (check this using paths), it is easily seen that (8) is equivalent to

$$\sigma_t'(m) \geq (\sigma_t'(m-k) + \sigma_t'(m+k))/2. \tag{9}$$

There is an obvious symmetry in the graphical construction which implies that

$$\begin{aligned}\sigma_t'(y) &= P(\text{there exists a } [0, t]\text{-path from } y \text{ to } \mathbf{Z} \mid \mathscr{A}) \\ &= P(\text{there exists a } [0, t]\text{-path from } \mathbf{Z} \text{ to } y \mid \mathscr{A}'),\end{aligned}$$

where $\mathscr{A}' = \{A \cap \xi_t^{\mathbf{Z}} = \varnothing\}$. Thus, $\sigma_t'(y)$ equals a kind of conditional population profile when the initial state is $\mathbf{Z}$, conditioned on the event that all the sites in A are vacant at time t. The vacancy of sites in A means the presence of random forbidden zones on either side of any $[0, t]$-path from $\mathbf{Z}$ to a site $y \in I$ (see Section 4). Thus, (9) is a statement about the shape of the conditional population profile σ_t' between two forbidden zones after symmetrization about the center of I. We have already shown how to prove a statement about the shape of a symmetric (unconditional) population profile, namely Theorem 3. In fact, the proofs of Theorem 3 and (9) are very similar, except that in the proof of (9), Theorem 2 must be used to handle the random forbidden zones. Details will appear in a future article.

We have seen that certain special properties of extremal paths are quite useful for deriving information about the behavior of the contact process. It is hoped that this report will arouse some interest in further investigations along these lines. In my opinion, there is much to be done. Rumors of the demise of the contact process are greatly exaggerated!

BIBLIOGRAPHY

1. C. Bezuidenhout and G. Grimmett, *The critical contact process dies out*, Ann. Probab., to appear.

2. R. Durrett, *The contact process*, 1974–1989 (appears elsewhere in this volume).

3. L. Gray, *Duality for general attractive spin systems, with applications in one dimension*, Ann. Probab. **14** (1985), 371–396.

4. D. Griffeath, *Additive and cancellative interacting particle systems*, Lecture Notes in Math. **724** (1979).

5. T. Harris, *Contact interactions on a lattice*, Ann. Probab. **2** (1974), 969–988.

6. ——, *A correlation inequality for Markov processes in partially ordered state spaces*, Ann. Probab. **5** (1977), 451–454.

7. ——, *Additive set-valued Markov processes and graphical methods*, Ann. Probab. **6** (1978), 355–378.

8. T. Liggett, *Interacting particle systems*, Springer, New York, 1985.

9. ——, *Nearest particle systems: Results and open problems*, in Stochastic Spatial Processes, Lecture Notes in Math. **1212** (1986).

SCHOOL OF MATHEMATICS, UNIVERSITY OF MINNESOTA
E-mail address: gray@ux.acs.umn.edu

Lectures in Applied Mathematics
Volume 27 (1991)

Limiting Behavior of a One-Dimensional System with Long Range Interactions

THOMAS M. LIGGETT

During the past twenty years, the study of interacting particle systems has concentrated on certain specific types of models. These have been chosen partly for their simplicity and mathematical appeal, partly because they exhibit phenomena such as phase transition which are of interest in other fields, and partly because they are well suited to analysis by certain important techniques. One type of model which has all of these attributes is known as a nearest particle system. This paper describes the progress which has been made during the past decade in understanding the limiting behavior of this process, with particular emphasis on the most tractable case in which it has the properties of attractiveness and reversibility.

The nearest particle systems we will describe are Markov processes η_t on the set of configurations $\eta \in \{0, 1\}^{\mathbf{Z}}$, where $\mathbf{Z}$ is the set of integers, which have infinitely many 1's in both directions. The state $\eta(k)$ at site k flips from a 1 to a 0 at rate one, and flips from a 0 to a 1 at rate $\beta(l, r)$, where l and r are the distances from k to the nearest sites to the left and right, respectively, at which there is a 1. The birth rate $\beta(l, r)$ is a strictly positive, bounded, symmetric function of l and r. The process has long range interactions, so some technical problems arise in its construction. These were resolved by Gray [3]. In this paper, we will assume that the process is attractive in the sense

---

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60K35.
The preparation of this paper was supported in part by NSF Grant DMS 86-01800.
This paper is in final form and no version of it will be submitted for publication elsewhere.

that the birth rate is an increasing function of the configuration, and so $\beta(l, r)$ is a decreasing function of l and r. In this case, the process can be extended as a Feller process to all of $\{0, 1\}^{\mathbf{Z}}$.

The attractiveness implies that if initially $\eta \equiv 1$, then the distribution of the process is monotonically decreasing in time. The process is said to *die out* if the limiting distribution υ as $t \to \infty$ is the pointmass on $\eta \equiv 0$, and is said to *survive* otherwise. In either case, υ is the largest invariant measure. The process is called *ergodic* if υ is the only invariant measure. It is said to be *supercritical* if there exists a $\lambda < 1$ so that the process with birth rates $\lambda\beta(l, r)$ survives. It is said to be *subcritical* if there exists a $\lambda > 1$ so that the process with birth rates $\lambda\beta(l, r)$ dies out. A process which is neither subcritical nor supercritical is called *critical*. The main problems which have been considered are:

(1) For what choices of birth rates does the process survive?
(2) When the process survives, are there invariant measures other than υ and the pointmass on $\eta \equiv 0$?
(3) What can be said about rates of convergence to υ?
(4) What differences in behavior are there between critical and supercritical processes?
(5) How can effective comparisons be made between different nearest particle systems, which enable one to show, for example, that if one survives, then the other does also?

One answer to the first question is provided by the following theorem. Part (ii) is based on work of Bramson and Gray [**2**]—see Chapter VII of Liggett [**6**] for the proof. Part (iii) is a recent result due to Bramson [**1**]. For many natural parametric families of birth rates, the theorem implies that the process survives for large values of the parameter and dies out for small values. For example, for the uniform birth process with rates

$$\beta(l, r) = \frac{\lambda}{l + r - 1},$$

it gives extinction for $\lambda \leq 1$ and survival for $\lambda > 4\log 2 \sim 2.77$. (This process gets its name from the fact that a total birth rate of λ is distributed uniformly over the sites between two successive ones.) (Note added in proof: T. Mountford has shown recently that this process survives for all $\lambda > 1$.)

THEOREM 1. (i) *If*

$$\sum_{l+r=n} \beta(l, r) \le 1$$

for all $n \ge 2$, *then the process dies out.*

(ii) *If*

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{l+r=n} \beta(l, r)[n \log n - l \log l - r \log r] > 2 \log 2,$$

then the process survives.

(iii) *For every* $\varepsilon > 0$, *there exists a choice of birth rates satisfying*

$$\sum_{l+r=n} \beta(l, r) = 1 + \varepsilon$$

for $n \ge 2$ *for which the process survives.*

The proof of the first part of the theorem is quite easy. It is based on the observation that blocks between successive 1's in the configuration contribute a net death rate, since each 1 has a death rate of one, and each block of 0's has a total birth rate of at most one:

$$\ulcorner 1\,0\,0\,0 \,|\, 1\,0 \,|\, 1 \,|\, 1\,0\,0\,0\,0 \,|\, 1 \,|\, 1 \,|\, 1\,0\,0\,0\,0 \,|\, 1$$

For the second part, one defines

$$h(\mu) = \sum_{n=1}^{\infty} \mu\{\eta\colon \eta(0) = \eta(n) = 1,\ \eta(k) = 0 \text{ for } 0 < k < n\} n \log n$$

for shift invariant μ, and then proves that there exist positive constants ε and K so that

$$\frac{d}{dt} h(\mu_t) \le -\varepsilon + K\mu_t\{\eta\colon \eta(0) = 1\},$$

where μ_t is the distribution at time t. The attractiveness assumption implies that $h(\mu_t)$ is nondecreasing in t. Therefore the density of ones is bounded below, so the process survives. Bramson's example for part (iii) is a truncated version of the uniform birth process—the birth rate is distributed uniformly over those sites which are within a fixed distance of a one. His result suggests that the uniform birth process survives for all $\lambda > 1$. However, it appears to be difficult to make a comparison which would lead to a proof that the uniform birth process survives if its truncated version does.

Since η_t has been extended as a Feller process to all configurations, one can start it with the configuration $\eta \equiv 0$. One might think that whenever

$$\lim_{l, r \to \infty} \beta(l, r) = 0,$$

the pointmass δ_0 on $\eta \equiv 0$ is an invariant measure. This is not necessarily the case, as can be seen from the following result proved in Liggett [5]. Intuitively, one should think that there are fixed 1's at $-\infty$ and $+\infty$, which may generate 1's in $\mathbf{Z}$ at positive times, in much the same way as Markov chains can enter the state space from boundary states.

THEOREM 2. *Let*

$$\alpha(n) = \sum_{l+r=n} (l \wedge r)\beta(l, r) \quad \text{and} \quad \alpha^*(n) = \max\{\alpha(k): k \le n\}.$$

Then $\eta \equiv 0$ *is a trap if*

$$\sum_{n \ge 2} \frac{1}{\alpha^*(n)} = \infty,$$

and is not a trap if

$$\sum_{n \ge 2} \frac{1}{\alpha(n)} < \infty.$$

Theorems 1 and 2 give about all that is known about general nearest particle systems. Much more is known about those for which $\upsilon \neq \delta_0$ is reversible. The starting point for this development is the following result due to Spitzer [**10**], which initiated the study of nearest particle systems.

THEOREM 3. *There exists a probability measure concentrating on configurations with infinitely many* 1*'s which is reversible for the process if and only if there exists a positive probability measure* $\beta(n)$ *with finite mean on the positive integers so that*

$$\beta(l, r) = \frac{\beta(l)\beta(r)}{\beta(l+r)} \tag{1}$$

for $l, r \ge 1$. *In this case, the reversible measure is the distribution of the stationary renewal process with interarrival times distributed according to* $\beta(n)$. *It is the largest invariant measure* υ.

An easy corollary of Spitzer's theorem is that if $\beta(l, r)$ can be written in the form (1) in terms of a positive function $\beta(n)$, but cannot be rewritten in that form in terms of a probability measure with finite mean, then the process dies out. For example, if

$$\beta(l, r) = c\left(\frac{1}{l} + \frac{1}{r}\right)^p = c\frac{l^{-p}r^{-p}}{(l+r)^{-p}} \tag{2}$$

for some positive parameters c and p, then (a) if $p \leq 1$, the process survives for all c, (b) if $1 < p \leq 2$, the process survives if and only if $c > [\sum_{n\geq 1} n^{-p}]^{-1}$, and (c) if $p > 2$, the process survives if and only if $c \geq [\sum_{n\geq 1} n^{-p}]^{-1}$. In this example, Theorem 2 implies that δ_0 is a trap if and only if $p \geq 1$. Note that there are critical processes which survive, and critical processes which die out.

In the remainder of the paper, we will assume that the birth rates have the form (1) for a probability density $\beta(n)$ with finite mean M. The attractiveness assumption translates into the logconvexity of this density. This implies that

$$\rho = \lim_{n\to\infty} \frac{\beta(n+1)}{\beta(n)} \leq 1$$

exists. An immediate consequence of Spitzer's theorem is that the process is critical if and only if $\rho = 1$. The renewal measure corresponding to $\beta(n)$ will always be denoted by υ.

A second result proved in Liggett [5] rules out the possibility of the existence of invariant measures other than υ under some additional conditions. The proof is based on the relative entropy technique (see, for example, Chapter II, Section 4 of Liggett [6]). When combined with Theorem 2, it implies that the process with birth rates given in (2) is ergodic if $p < 1$. To see this, it is enough to note that the limiting distribution when initially $\eta \equiv 0$ is both invariant and translation invariant, and hence must be υ.

THEOREM 4. *Suppose that the process is supercritical, or is critical and satisfies*

$$\sum_{n=1}^{\infty} \frac{\beta^2(n)}{\beta(2n)} < \infty. \tag{3}$$

Then υ is the only measure on the set of configurations with infinitely many 1's which is both translation invariant and invariant for the process.

Our final topic is rates of convergence to υ of the distribution of the process as $t \to \infty$. One reason for being interested in this is that we hope to see essentially different rates in supercritical and critical cases.

The first step is to decide how to measure the rate of convergence. When δ_0 is invariant, the convergence cannot be uniform in the initial configuration. That suggests that an L_2 type of convergence is appropriate. Let $S(t)$ be the $L_2(\upsilon)$ semigroup for the process, and $\|\cdot\|$ be the norm in that space. The process is said to converge exponentially rapidly in $L_2(\upsilon)$ if there exists an $\varepsilon > 0$ so that

$$\left\|S(t)f - \int f\, d\upsilon\right\| \le e^{-\varepsilon t}\left\|f - \int f\, d\upsilon\right\|$$

for all $f \in L_2(\upsilon)$. We will take ε to be the largest number with this property.

Before stating conditions under which the nearest particle system has this property of exponential convergence, we illustrate its meaning in a simpler context.

EXAMPLE. Consider a positive recurrent birth and death chain on the nonnegative integers with stationary distribution π, whose birth and death rates are bounded above and below by positive constants. Then the chain converges exponentially rapidly in $L_2(\pi)$ if and only if π decays exponentially, in the sense that there is a constant C so that

$$\sum_{k=n}^{\infty} \pi(k) \le C\pi(n).$$

Versions of this result can be found in Sullivan [**11**], Lawler and Sokal [**4**], and Liggett [**7**].

The following theorem is obtained by combining results from Liggett, [**7**] and [**8**]. It not only gives a necessary and sufficient condition for exponential convergence, but under additional regularity assumptions also provides bounds on ε which are good enough to imply (under a second moment assumption) that its critical exponent is 2.

THEOREM 5. *Suppose that* (3) *holds.*

(i) *The process converges exponentially rapidly in* $L_2(\upsilon)$ *if and only if it is supercritical.*

(ii) *If* $\beta(n+m)$ *is totally positive of order three, then*

$$\varepsilon \ge \frac{\beta(1)}{4M}(1-\rho)^2.$$

(iii)
$$\varepsilon \le 4(1-\rho)^2 \sum_{n=1}^{\infty} n^2 \beta(n) \rho^{-n}.$$

REMARK. The logconvexity of $\beta(n)$ which we have been assuming is equivalent to $\beta(n+m)$ being totally positive of order two. Thus we need assume only slightly more regularity of the same type to get the lower bound on ε. This assumption is satisfied by the examples in (2).

A natural question suggested by the above theorem is whether some type of algebraic convergence in $L_2(\upsilon)$ occurs in the critical case. Here our results are less complete. The proofs are contained in Liggett [9].

There is some subtlety in deciding what form the bound should take, since if $\|S(t)f - \int f\,d\upsilon\| \le \|f - \int f\,d\upsilon\| t^{-\alpha}$ for all f and some $\alpha > 0$, it would follow that the convergence is actually exponentially fast. Therefore the norm on the right is usually replaced by some type of Lipschitz norm. Before discussing nearest particle systems, we return to the birth and death chain setting.

EXAMPLE (CONTINUED). Suppose that the transition rates of the birth and death chain are bounded above and below by positive constants, and satisfy

$$0 \le q(k, k-1) - q(k, k+1) \le C/k$$

for $k \ge 1$ and some constant C. Suppose also that the stationary distribution satisfies

$$\sum_{k \ge n} \pi(k) \le C n \pi(n)$$

for some C. Fix a $q > 1$. If

$$\sum_{k \ge 0} k^{\alpha} \pi(k) < \infty$$

for some $\alpha > 2q$, then there is a constant C so that

$$\left\| S(t)f - \sum f(k)\pi(k) \right\|^2 \le C \frac{\sup_k |f(k+1) - f(k)|^2}{t^{q-1}}$$

for all f. Conversely, if this is the case, then

$$\sum_{k \ge 0} k^{\alpha} \pi(k) < \infty$$

for all $\alpha < 2q$.

For the statement of the next result, define

$$|||f||| = \sum_k \sup_{\eta} |f(\eta_k) - f(\eta)|$$

for continuous f on $\{0, 1\}^{\mathbf{Z}}$, where η_k is the configuration obtained from η by flipping the kth coordinate. It is a type of Lipschitz norm which has been useful often in interacting particle systems.

THEOREM 6. *Suppose that* $\beta(m+n)$ *is totally positive of order three, and that* $\beta(n)$ *is regularly varying. Fix a* $q > 1$.

(i) *Suppose that*

$$\sum_{k\geq 0} k^{2q+2}\beta(k) < \infty.$$

Then there exists a constant C *so that*

$$\left\|S(t)f - \int f\,dv\right\|^2 \leq C\frac{|||f|||^2}{t^{q-1}} \tag{4}$$

for all $f \in L_2(v)$.

(ii) *Suppose that there exists a constant* C *so that* (4) *holds for all* $f \in L_2(v)$. *Then*

$$\sum_{k\geq 0} k^{\alpha}\beta(k) < \infty$$

for all $\alpha < q - 2$.

REMARKS. (i) For the sake of simplicity, Theorems 5 and 6 and the results about birth and death chains have not been stated in maximal generality. For the more general statements, see Liggett [7], [9].

(ii) It would be interesting to narrow the gap in the moments appearing in the two parts of Theorem 6.

Finally, we make a few comments about the proofs of Theorems 5 and 6. The first step in each case is to reduce the problem to proving a statement about the generator Ω of the process. For functions f in its domain $D(\Omega)$, define the Dirichlet form by $\mathbf{E}(f, f) = -\int f\Omega f\,dv$. Then exponential L_2 convergence is equivalent to the existence of a constant C so that

$$\left\|f - \int f\,dv\right\|_2^2 \leq C\mathbf{E}(f, f)$$

for all f. The corresponding criterion for algebraic convergence is more complicated. Suppose that $V(f)$ is a quadratic functional of f which is not changed by the addition of a constant to f and which satisfies $V(S(t)f) \leq V(f)$ for all f and $t \geq 0$. Fix $1 < q < \infty$, and let $p^{-1} + q^{-1} = 1$. If the process is reversible with respect to v, then the following statements are equivalent.

(a) There exists a constant C so that

$$\left\|f - \int f\,dv\right\|_2^2 \leq C[\mathbf{E}(f, f)]^{1/p}[V(f)]^{1/q} \quad \text{for all } f \in D(\Omega).$$

(b) There exists a constant C so that

$$\left\|S(t)f - \int f\,d\upsilon\right\|_2^2 \leq C\frac{V(f)}{t^{q-1}} \quad \text{for all } f \in L_2(\upsilon).$$

In order to verify the appropriate statement about the Dirichlet form, we use a special construction of the renewal measure υ in terms of a product measure. To describe it, let $g(n)$ be the renewal sequence associated with the density $\beta(n)$. The logconvexity of $\beta(n)$ implies the logconvexity of $g(n)$, as was proved by de Bruijn and Erdös in 1953. We can therefore define a probability measure $\pi(n)$ on the nonnegative integers by $\pi(0) = g(1)$ and

$$\pi(n) = \frac{g(n+1)}{g(n)} - \frac{g(n)}{g(n-1)} \quad \text{for } n \geq 1.$$

Let μ be the product measure on $\{0, 1, 2, \dots\}^{\mathbf{Z}}$ with marginals π on each coordinate. Define a mapping $T: \{0, 1, 2, \dots\}^{\mathbf{Z}} \to \{0, 1\}^{\mathbf{Z}}$ by $T(X) = \eta$, where

$$\eta(n) = 1 \Leftrightarrow X(n+k) \leq k \quad \text{for all } k \geq 0.$$

It turns out that the image of μ under T is the renewal measure υ. This makes it possible to write the variance of f with respect to υ as the variance of a function of i.i.d. random variables with distribution π, and then to carry out the estimates needed for the sufficiency parts of Theorems 5 and 6. These estimates require rather detailed information about the behavior of the renewal sequence $g(n)$. The proofs of the necessary parts involve the computation of the variance of f, $\mathbf{E}(f, f)$, and $|||f|||$ for carefully chosen functions f.

References

1. M. Bramson, *Survival of nearest particle systems with low birth rate*, Ann. Probab. **17** (1989), 433–443.

2. M. Bramson and L. Gray, *A note on the survival of the long-range contact process*, Ann. Probab. **9** (1981), 885–890.

3. L. Gray, *Controlled spin-flip systems*, Ann. Probab. **6** (1978), 953–974.

4. G. F. Lawler and A. D. Sokal, *Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality*, Trans. Amer. Math. Soc. **309** (1988), 557–580.

5. T. M. Liggett, *Attractive nearest particle systems*, Ann. Probab. **11** (1983), 16–33.

6. ——, *Interacting Particle Systems*, Springer, New York, 1985.

7. ——, *Exponential L_2 convergence of attractive reversible nearest particle systems*, Ann. Probab. **17** (1989), 403–432.

8. ——, *Total positivity and renewal theory*, in Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin, Academic Press, 1989, 141–162.

9. ——, *L_2 rates of convergence for attractive reversible nearest particle systems: the critical case*, Ann. Probab., to appear (1991).

10. F. Spitzer, *Stochastic time evolution of one-dimensional infinite particle systems*, Bull. Amer. Math. Soc. **83** (1977), 880–890.

11. W. G. Sullivan, *The L^2 spectral gap of certain positive recurrent Markov chains and jump processes on Z*, Z. Wahrsch. verw. Gebiete **67** (1984), 387–398.

UNIVERSITY OF CALIFORNIA, LOS ANGELES

E-mail address: TML@MATH.UCLA.EDU

Lectures in Applied Mathematics
Volume 27 (1991)

Topics in Percolation

CHARLES M. NEWMAN

Abstract. We introduce percolation models and discuss such matters as dynamic renormalization and the continuity of the percolation transition, uniqueness of infinite clusters and the layering transition, invasion percolation and the trapping transition.

1. Introduction. The purpose of this paper is to give a brief (and hence rather incomplete) survey of a number of loosely related topics from percolation theory. A good general account of (independent) percolation theory is to be found in [G]. In this section we introduce percolation models and then discuss the problem of continuity of the percolation transition. Several recent results about this matter rely on a dynamic renormalization technique which we describe in Section 2. In Section 3 we focus on the issue of uniqueness of the infinite cluster emphasizing the case of percolation models which are dependent or which involve unusual branching lattices; the latter case gives rise to a "layering" transition distinct from the usual percolation transition. Yet another transition plays the starring role in Section 4; this one involves "trapping" and is related to the disappearance of the "external surface" of the infinite cluster. A major issue here is the relation between this transition and invasion percolation, a certain dynamic growth model.

Nearest-neighbor bond percolation on $\mathbf{Z}^d$ is described by 0- or 1-valued random variables $\{n_b : b \in \mathscr{B}_d\}$. A bond b is said to be open (or occupied or conducting or $\cdots$) when $n_b = 1$ and closed (or vacant

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60K35, 82A43.
Research supported in part by NSF Grants DMS-851483 and DMS-8902156, by AFOSR Grant 88-0189 and by AFOSR Contract F49620-86-C0130 to the Arizona Center for Mathematical Sciences under the U.R.I. Program.

0075-8485/91 $1.00 + $.25 per page

or insulating or $\cdots$) when $n_b = 0$. Here $\mathscr{B}_d$ is the set of pairs $\{x, y\}$ of sites in $\mathbf{Z}^d$ with $\|x - y\| = 1$, where $\|\cdot\|$ denotes Euclidean length. The random graph with vertex set $\mathbf{Z}^d$ and edge set consisting of all open bonds decomposes into *clusters* (i.e., maximal connected components) and *percolation* is said to occur when some cluster is infinite.

A percolation model is determined by $\mathbf{P}$, the joint probability distribution of the n_bs. We denote by $\mathbf{P}_p$ the distribution for the *independent model with bond density* p, corresponding to independent n_bs with $\mathbf{P}_p(n_b = 1) = p$ for each b [**BH**]. We will also occasionally mention the *independent site percolation model with site density* λ where sites are independently occupied (with probability λ) or vacant; here the vertex set of the random graph consists of all occupied sites and the edge set consists of all bonds in $\mathscr{B}_d$ between pairs of occupied sites. Theorem 1, from [**H1**, **H2**], establishes the existence of a percolation phase transition.

THEOREM 1. *Define* $\theta(p) = \mathbf{P}_p$ *(the cluster of the origin is infinite). For* $d \geq 2$, *there is a critical density* $p_c = p_c(d)$ *in* $(0, 1)$ *such that*

$$p < p_c \quad \textit{implies } \theta(p) = 0,$$

and hence no percolation occurs with probability one (w.p.1) while

$$p > p_c \quad \textit{implies } \theta(p) > 0$$

and hence percolation occurs (w.p.1).

The status of (non)percolation at the critical point will be discussed following the next two theorems which concern behavior away from the critical point. Theorem 2 is about the subcritical regime and combines results of [**H1**] and [**M**, **AB**].

THEOREM 2. *Define* $\tau_p(x, y) = \mathbf{P}_p$ *(x and y are in the same cluster); then*

$$p < p_c \quad \textit{implies } \tau_p(0, x) = O(\exp(-m_p\|x\|))$$

as $\|x\| \to \infty$ *for some* $m_p > 0$.

Theorem 3 is about the supercritical regime and combines results of [**CCN2**] and [**GM**]; the latter result will be discussed in Section 2 below.

THEOREM 3. *Define* $\tau'_p(x, y) = \mathbf{P}_p$ *(x and y are in the same finite cluster); then*

$$p > p_c \quad \textit{implies } \tau'_p(0, x) = O(\exp(-m'_p\|x\|))$$

as $\|x\| \to \infty$ *for some* $m'_p > 0$.

To see why the issue of (non)percolation at p_c is important, we first note that θ is necessarily a right-continuous function of p. This is so because it is a nondecreasing function of p and also is the decreasing limit as $L \to \infty$ of the continuous functions

$$\theta_L(p) = \mathbf{P}_p \text{ (there is a path of open bonds from 0 to some } x \text{ with } \|x\| = L).$$

Theorem 4 then shows that $\theta(p)$ is continuous for all p if and only if $\theta(p_c) = 0$; in this case the percolation phase transition is said to be continuous. Theorem 4 combines results of **[BeK]** and **[AKN]**; the latter result is Theorem 8 in Section 3 below.

THEOREM 4. *For each $p > p_c$, θ is left-continuous (and hence continuous) at p.*

The next two theorems sum up the current status of (non)percolation at p_c. The $d = 2$ part of Theorem 5 combines results of **[Ha1]** and **[K]** while the large d part is from **[HS1, HS2]**.

THEOREM 5. (i) *For $d = 2$, $\theta(p_c) = 0$.*
(ii) *There is a finite integer $d_0 \geq 6$ such that $d > d_0$, $\theta(p_c) = 0$.*

We remark that the results of **[Ha1, K]** are rather special to bond percolation; for a proof of nonpercolation at the critical point sufficiently general to cover, say, independent site percolation on $\mathbf{Z}^2$, see **[R]**. We also remark that for nearest-neighbor percolation, the best current value of d_0 seems to be 92. On the other hand, for other models with spread-out but still finite range bonds, $d_0 = 6$ **[HS1, HS2]**. Theorem 6, which is based on "dynamic renormalization" methods as discussed in Section 2 below, is a result of **[BGN1]**.

THEOREM 6. *Set $d \geq 3$.* (i) *For independent nearest-neighbor percolation on the half-space (in which $\mathbf{Z}^d$ is replaced by $\mathbf{Z}^{d-1} \times \mathbf{Z}_+$), there is no percolation at the critical point (w.p.1).*
(ii) *A resolution of the following open problem would imply that $\theta(p_c) = 0$ (in $\mathbf{Z}^d$):*

> *Prove that percolation in $\mathbf{Z}^d$ at density p implies percolation in $\mathbf{Z}^{d-1} \times \mathbf{Z}_+$ at the same density p.*

In the next section, we discuss the dynamic renormalization technique introduced in **[BGN1]**. Meanwhile we remark that a variation of these techniques has led to a proof for $d \geq 2$ that the contact process on $\mathbf{Z}^{d-1}$ dies out at its critical point [**BG**]. The contact process has a representation as directed percolation in $\mathbf{Z}^{d-1} \times T$ [**Ha2**] where the time axis T can be either continuous ($[0, \infty)$) or discrete ($\mathbf{Z}_+$); hence, the geometric structure of this process is very close to that of ordinary percolation in a half-space.

2. Dynamic renormalization. The basic technical fact which leads to Theorem 6 is a proof of the following [**BGN1**]:

> Percolation in $\mathbf{Z}^{d-1} \times \mathbf{Z}_+$ at density p implies percolation in the quarter slice $Q_L = \{-L, \dots, L\}^{d-2} \times \mathbf{Z}_+^2$ at density $p - \varepsilon$ for some $\varepsilon > 0$ and $L < \infty$.

This fact also leads to the conclusion that the half-space critical density equals the limit as $L \to \infty$ of the Q_L critical densities. In [**GM**] new ideas were added to the dynamic renormalization techniques of [**BGN1**] to prove the following technical fact:

> Percolation in $\mathbf{Z}^d$ at density p implies percolation in Q_L at density $p + \varepsilon$ for any $\varepsilon > 0$ and some $L = L(\varepsilon) < \infty$.

This fact implies that the limit of Q_L critical densities equals the full-space critical density p_c, which is a basic ingredient in obtaining results on supercritical behavior such as Theorem 3 above. We note that the open problem presented in part (ii) of Theorem 6 is basically to show that the ε of [**GM**] can be set to zero.

In the remainder of this section we discuss a bit about the dynamic renormalization arguments of [**BGN1**] and how they differ from static renormalization methods such as those of [**R, ACCFR**]. Renormalization arguments with dynamic aspects have appeared previously in the contexts of long range percolation [**AiN**] and directed percolation [**BDS, D**]. For a more complete discussion than that given here, but still without the many technical complications of [**BGN1**], see [**BGN2**] for a proof of a simplified version of Theorem 6 (in which the half-space is replaced by the orthant $\mathbf{Z}_+^d$).

How are renormalization methods used to prove that there is percolation in Q_L at density p' (where $p' = p - \varepsilon$ or $p + \varepsilon$ above)? First partition Q_L into cubes which are translates of $\{-L, \dots, L\}^d$ in the obvious way and think of each cube as a "renormalized site" in the renormalized lattice $\mathbf{Z}_+^2$. Say that a renormalized site is occupied if the corresponding cube is "well connected" in some appropriate sense. Show that the occupied renormalized sites percolate in $\mathbf{Z}_+^2$ and then conclude that the original model percolates in Q_L.

In static renormalization. The meaning of "well connected" is typically the same (up to translations) for each renormalized site, and percolation in $\mathbf{Z}_+^2$ is shown by testing the occupation status of all cubes simultaneously.

In dynamic renormalization. Both the definition of "well connected" and the testing of cube status are done dynamically, one cube at a time, and the meaning of "well connected" for a given cube depends on the microstate of neighbor cubes. A cube is not tested until (at least) one of its neighbors is found to be occupied; as in a standard algorithm for the computer simulation of the cluster of the origin for site percolation in $\mathbf{Z}_+^2$, the cluster (and its boundary) is "grown", one (renormalized) site at a time. To show percolation in the renormalized lattice, it suffices if

$$\mathbf{P}\ (\text{a cube is occupied} \mid \text{it is tested}) > \lambda_c(\mathbf{Z}_+^2) + \varepsilon'$$

where $\lambda_c(\mathbf{Z}_+^2)$ is the critical density for independent nearest-neighbor site percolation in $\mathbf{Z}_+^2$.

The dynamic renormalization procedure just described starts with, say, an independent bond model, renormalizes it into a dependent site model and then compares the latter to an independent site model to prove percolation in the original model. In order to point out a certain geometric feature of this type of procedure which was not previously emphasized, namely, that it has an underlying tree-like structure, we apply a variant of the argument to dependent bond percolation to obtain Theorem 7, a result which is implicit in **[BGN1, BGN2]**.

THEOREM 7. *For a (dependent) nearest-neighbor bond percolation model, $\{n_b : b \in \mathscr{B}_d\}$, define M_x to be 1 if x belongs to the cluster of the origin and otherwise to be zero. Then the random field $\{M_x : x \in \mathbf{Z}^d\}$ stochastically dominates $\{M'_x(\lambda_0) : x \in \mathbf{Z}^d\}$ for a "renormalized site density" $\lambda_0 \in [0, 1]$ defined below. Here, for a given λ, $M'_x(\lambda)$ is 1 if x belongs to the cluster of the origin in an independent nearest-neighbor site percolation model on $\mathbf{Z}^d$ at density λ and otherwise $M'_x(\lambda) = 0$. Thus if $\lambda_0 > \lambda_c(\mathbf{Z}^d)$, the critical site density in $\mathbf{Z}^d$, then the original bond model percolates (with positive probability). To define λ_0, first denote by $\mathscr{T}$ the collection of finite subtrees t of the graph $(\mathbf{Z}^d, \mathscr{B}_d)$*

which contain the origin and then for each $t \in \mathcal{T}$ *denote by* $\mathcal{U}_t$ *the collection of nonempty sets* u *of bonds* $b \in \mathcal{B}_d$ *such that*

(a) *each* b *in* u *touches exactly one vertex in* t, *and*

(b) t *adjoined with the bonds (and new vertices) of* u *is still a tree.*

Then λ_0 *is defined as*

$$\lambda_0 = \inf_{t\in\mathcal{T}} \inf_{u\in\mathcal{U}_t} \inf_{b\in u} \mathbf{P}(n_b = 1 \mid n_{b'} = 1 \text{ for } b' \in t \text{ and } n_{b''} = 0 \text{ for } b'' \in u\setminus\{b\}).$$

3. Uniqueness of infinite clusters. When percolation occurs, is the infinite cluster unique? Theorem 8, from [**AKN**], gives a positive answer, at least for independent percolation on $\mathbf{Z}^d$.

THEOREM 8. *If* $\theta(p) > 0$, *then*

$$\mathbf{P}_p\,(\textit{there is exactly one infinite cluster}) = 1.$$

In the rest of this section we discuss the status of uniqueness for more general percolation models—first for dependent models on $\mathbf{Z}^d$ and then for models on more exotic lattices than $\mathbf{Z}^d$. Dependent percolation models are of some importance in statistical mechanics. There are for example (at least) two ways in which dependent percolation and Ising models are related. Ising models are certain (dependent) ± 1-valued random fields $\{S_x : x \in \mathbf{Z}^d\}$ and one may obtain a dependent site percolation model by declaring x to be occupied if $S_x = +1$ and vacant if $S_x = -1$. For $d = 2$ (but not larger d) there is a close connection between the percolation transition in this dependent site model and the usual Ising phase transition, as shown in [**CNPR**]. (Some recent applications of this relation may be found in [**AN1, AN2**] and [**NSt**].) There is another way of relating dependent percolation to Ising models which is more subtle but provides a connection valid for any d between percolation and (ferromagnetic) Ising transitions. Here one deals with *bond* percolation models known as random cluster or FK models [**FK**]. General discussions of FK models may be found in [**ACCN**] (covering the ferromagnetic context) and in [**N**] (covering the nonferromagnetic context such as occurs in Ising spin glasses).

The next example shows that uniqueness can fail for a dependent percolation model $\{n_b : b \in \mathcal{B}_d\}$ even though it is translation invariant (with respect to $\mathbf{Z}^d$-translations) and is "almost independent" in the sense that $\{n_b : b \in V\}$ and $\{n_b : b \in V'\}$ are independent of each other whenever the two sets of bonds V and V' have no site in common.

EXAMPLE. Let $\{S_x : x \in \mathbf{Z}^3\}$ be independent symmetric ± 1-valued random variables. Define $n_{\{x,y\}} = 1$ if $S_x = S_y$ and 0 otherwise and let $\mathbf{P}$ denote the distribution of $\{n_b : b \in \mathscr{B}_3\}$; then

$$\mathbf{P}\,(\text{there are exactly two infinite clusters}) = 1\,.$$

This follows from the fact that the sites x with $S_x = +1$ (respectively -1) form an independent site percolation model with density $1/2$ and the fact that the critical density $\lambda_c(\mathbf{Z}^3)$ is strictly less than $1/2$ [**CR**]. Thus there is one infinite $\{n_b\}$-cluster coming from each of the plus and minus site clusters.

In spite of this example, Theorem 9, from [**BuK**] (with a minor improvement from [**GKN**]), gives a nice general result guaranteeing uniqueness for certain dependent models.

THEOREM 9. *Assume that the distribution* $\mathbf{P}$ *of a percolation model* $\{n_b : b \in \mathscr{B}_d\}$ *is translation invariant and for each* b, *its conditional probability with respect to the* σ*-field generated by* $\{n_{b'} : b' \in \mathscr{B}_d \setminus \{b\}\}$ *satisfies*

$$\mathbf{P}(n_b = 1 | \{n_{b'} : b' \neq b\}) > 0 \qquad w.p.1\,.$$

Then

$$\mathbf{P}\,(\textit{there is more than one infinite cluster}) = 0\,.$$

We remark that Theorem 9 is applicable to (ferromagnetic) FK models and leave it to the reader to decide why it is not applicable to the example given above. For an application of Theorem 9 to spin glasses and for extensions of Theorem 9 (e.g., to long-range bond models), see [**GKN**].

The proofs of Theorems 8 and 9 rely on the fact that the lattice $\mathbf{Z}^d$ may be approximated by finite regions with much less surface than volume. What happens to uniqueness in lattices where this property fails? A simple example is the infinite homogeneous tree $\mathbf{T}_k$ with coordinate number $k+1 \geq 3$ for each site. Here it is clear, because of the absence of closed loops in this graph, that when percolation occurs, there must be infinitely many infinite clusters. But what happens in a lattice like $\mathbf{T}_k \times \mathbf{Z}$ where there are closed loops but also too much surface? If the tree $\mathbf{T}_k$ is thought of as a "branching line", then $\mathbf{T}_k \times \mathbf{Z}$ is a "branching plane" which possesses some of the features of the tree $\mathbf{T}_k$ and some of the features of the plane $\mathbf{Z}^2$. One may ask: which features dominate in the issue of uniqueness? Theorem 10, from [**GN**], shows that the answer depends on the bond density p: there is a "layering transition" in which the model changes from tree-like to plane-like behavior.

THEOREM 10. *Consider independent nearest-neighbor bond percolation on* $\mathbf{T}_k \times \mathbf{Z}$ *with bond density* p. *If* $k \geq 2$ *and* p *is close to* 1, *there is exactly one infinite cluster* (*which intersects each* **Z**-*line in a positive density of sites*) w.p.1. *If* k *is sufficiently large, then there is an open interval of values of* p *such that there are infinitely many infinite clusters* (*each of which intersects each* **Z**-*line in only finitely many sites*) w.p.1.

We note that the tree-like behavior is shown in [GN] to occur for all $k \geq 2$ providing different bond densities are allowed for the $\mathbf{T}_k$-bonds and the **Z**-bonds. Extensions of Theorem 10 which demonstrate the existence of a layering transition for Ising (and FK) models may be found in [NW]. The analogue of Theorem 10 for directed percolation (alias, the contact process) appears in [**P**].

4. The trapping transition. Invasion Percolation (IP) is a stochastic growth model for the microscopic development of a two-fluid (say, water and oil) interface in a porous medium [**LB**, **CKLW**]. The (static) random medium is described by i.i.d. continuous random variables $\{W_b : b \in \mathscr{B}_d\}$; W_b represents the difficulty of water displacing ("invading") the oil originally contained in bond b. In the (dynamic) invasion process, water invades one new bond at a time by choosing the least difficult one on the boundary of the currently invaded region. Invasion percolation with trapping is a variation of this growth model with the extra rule that any finite region of oil which is trapped by water (i.e., which is such that every path from the region to infinity must pass through currently invaded bonds) can no longer be invaded.

Simulation studies of IP and of IP with trapping were carried out in [**WW**, **W**]. It was found that IP had large scale fractal behavior which was closely related to the critical behavior of the usual percolation model (see Section 1 above) at its critical density p_c; some of these discoveries were later rigorously verified in [**CCN1**]. In simulating IP with trapping in [**W**, **WW**], the invasion process was allowed to proceed in a fixed 2- or 3-dimensional simulation region until all bonds were either invaded or trapped. Preliminary conclusions were that for $d = 2$, the large scale fractal behavior was again determined by the usual percolation model critical behavior but for $d = 3$, a "new universality class" arose. One ingredient in these conclusions was the perceived relation between (non)trapping and the percolation of vacant bonds in the usual percolation model. From this perspective $d = 2$ is degenerate because the critical density $1-p_c$ for vacant bond percolation equals p_c (equals $1/2$) in that dimension [K].

The analysis of IP with trapping of [W, WW] was subsequently shown to be flawed and their preliminary conclusions were cast in doubt by subsequent work of [CCN3], which showed that IP with trapping was related to a "trapping transition" in the usual percolation model occurring at a density κ_c which exceeds $1-p_c$ even for $d = 2$. The trapping transition, as defined by Theorem 11, is related to the disappearance of the "external surface" of the infinite cluster (see, e.g., [NS]). The $d = 2$ case of Theorem 11 was proved in [CCN3] and the general result is from [AG].

THEOREM 11. *For a bond percolation model* $\{n_b: b \in \mathscr{B}_d\}$, *define the random graph* G *with vertex set* $\mathbf{Z}^d$ *and edge set* $\mathscr{B}_d \backslash \{b: b$ *belongs to an infinite cluster of occupied bonds*$\}$. *For* $d \geq 2$, *there is a* $\kappa_c = \kappa_c(d)$ *in* $(0, 1)$ *with* $\kappa_c > 1 - p_c$ *such that*

$$p < \kappa_c \quad \text{implies } \mathbf{P}_p\ (G \text{ has an infinite component}) = 1,$$

$$p > \kappa_c \quad \text{implies } \mathbf{P}_p\ (G \text{ has an infinite component}) = 0.$$

The analysis of [CCN3] indicated two things regarding the fractal behavior of IP with trapping. First, that no new universality class should arise unless one already occurs in the trapping transition at κ_c. Second, that in this regard, $d = 2$ is not a degenerate case. Consequently, simulations were performed in [PNM] to test the null hypothesis that the $d = 2$ trapping transition has the same critical behavior as does the usual $d = 2$ percolation transition at $p_c = 1/2$. For example, if $\chi(p)$ denotes the mean (w.r.t. $\mathbf{P}_p$) size of the cluster of the origin and $\overline{\chi}(p)$ denotes the mean size of the "trap" of the origin (i.e., the maximal connected component in G which contains the origin), then the null hypothesis predicts equality of the critical exponents γ and $\overline{\gamma}$ defined by

$$\chi(p) \sim (p_c - p)^{-\gamma} \quad \text{as } p \uparrow p_c, \qquad \overline{\chi}(p) \sim (p - \kappa_c)^{-\overline{\gamma}} \quad \text{as } p \downarrow \kappa_c.$$

The simulations of [PNM] are completely consistent with the equality of these (and other) critical exponents; thus there appears to be no new universality class in IP with trapping. Incidentally, κ_c does not exceed $1-p_c$ by very much; for $d = 2$, $1-p_c = 1/2$ (exactly) while $\kappa_c \approx 0.520$ (numerically).

REFERENCES

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

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

[ACCN] M. Aizenman, J. T. Chayes, L. Chayes, and C. M. Newman, *Discontinuity of the order parameter in one-dimensional* $1/|x-y|^2$ *Ising and Potts models*, J. Statist. Phys. **50** (1988), 1–40.

[AG] M. Aizenman and G. R. Grimmett, in preparation.

[AKN] M. Aizenman, H. Kesten, and C. M. Newman, *Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation*, Comm. Math. Phys. **111** (1987), 505–532.

[AN1] D. B. Abraham and C. M. Newman, *Wetting in a three-dimensional system: an exact solution*, Phys. Rev. Lett. **61** (1988), 1969–1972.

[AN2] ——, *Surfaces and Peierls contours: 3-d wetting and 2-d Ising percolation*, Comm. Math. Phys. **125** (1989), 181–200.

[AiN] M. Aizenman and C. M. Newman, *Discontinuity of the percolation density in one dimensional* $1/|x-y|^2$ *percolation models*, Comm. Math. Phys. **107** (1986), 611–647.

[BDS] M. Bramson, R. Durrett, and G. Swindle, *Statistical mechanics of crabgrass*, Ann. Probab. **17** (1989), 444–481.

[BG] C. E. Bezuidenhout and G. R. Grimmett, *The critical contact process dies out*, Ann. Probab. **18** (1990), 1462–1482.

[BGN1] D. J. Barsky, G. R. Grimmett, and C. M. Newman, *Percolation in half-spaces: equality of critical densities and continuity of the percolation probability*, Probab. Theory Rel. Fields, to appear.

[BGN2] ——, *Dynamic renormalization and continuity of the percolation transition in orthants*, in Spatial Stochastic Processes (K. S. Alexander and J. Watkins, eds.), Birkhäuser, Boston, to appear.

[BH] S. R. Broadbent and J. M. Hammersley, *Percolation processes*, I. *Crystals and mazes*, Proc. Camb. Phil. Soc. **53** (1957), 629–641.

[BeK] J. van den Berg and M. Keane, *On the continuity of the percolation probability function*, in Conference on Modern Analysis and Probability (R. Beals et al., eds.), Contemp. Math. **26** (1984), 61–65.

[BuK] R. M. Burton and M. Keane, *Density and uniqueness in percolation*, Comm. Math. Phys. **121** (1989), 501–505.

[CCN1] J. T. Chayes, L. Chayes, and C. M. Newman, *The stochastic geometry of invasion percolation*, Comm. Math. Phys. **101** (1985), 383–407.

[CCN2] ——, *Bernoulli percolation above threshold: an invasion percolation analysis*, Ann. Probab. **15** (1987), 1272–1287.

[CCN3] ——, unpublished.

[CKLW] R. Chandler, J. Koplik, K. Lerman, and J. F. Willemsen, *Capillary displacement and percolation in porous media*, J. Fluid Mech. **119** (1982), 249–267.

[CNPR] A. Coniglio, C. R. Nappi, F. Peruggi, and L. Russo, *Percolation and phase transitions in the Ising model*, Comm. Math. Phys. **51** (1976), 315–323.

[CR] M. Campanino and L. Russo, *An upper bound for the critical probability for the three-dimensional cubic lattice*, Ann. Probab. **13** (1985), 478–491.

[D] R. Durrett, *A new method for proving the existence of a phase transition*, in Spatial Stochastic Processes (K. S. Alexander and J. Watkins, eds.), Birkhäuser, Boston, to appear.

[FK] C. M. Fortuin and P. W. Kasteleyn, *On the random cluster model* I. *Introduction and relation to other models*, Physica **57** (1972), 536–564.

[G] G. R. Grimmett, *Percolation*, Springer-Verlag, New York, 1989.

[GKN] A. Gandolfi, M. Keane, and C. M. Newman, *Uniqueness of the infinite component in a random graph with applications to percolation and spin glasses*, preprint (1989).

[GM] G. R. Grimmett and J. Marstrand, *The supercritical phase of percolation is well behaved*, Proc. Royal Soc. London Ser. A **430** (1990), 439–457.

[GN] G. R. Grimmett and C. M. Newman, *Percolation in infinity plus one dimensions*, in Disorder in Physical Systems (G. R. Grimmett and D. J. A. Welsh, eds.), Oxford Univ. Press, Oxford, 1990, pp. 167–190.

[H1] J. M. Hammersley, *Percolation processes. Lower bounds for the critical probability*, Ann. of Math. Statist. **28** (1957), 790–795.

[H2] ——, *Bornes supérieures de la probabilité critique dans un processus de filtration*, in Le Calcul des Probabilités et ses Applications, CNRS, Paris, 1959, pp. 17–37.

[Ha1] T. E. Harris, *A lower bound for the critical probability in a certain percolation process*, Proc. Camb. Phil. Soc. **56** (1960), 13–20.

[Ha2] ——, *Additive set-valued Markov processes and graphical methods*, Ann. Probab. **6** (1978), 355–378.

[HS1] T. Hara and G. Slade, *The mean-field critical behaviour of percolation in high dimensions*, in Proc. IXth International Congress on Mathematical Physics (B. Simon, A. Truman, and I. M. Davies, eds.), Adam Hilger, Bristol, 1989, pp. 450–453.

[HS2] ——, *Mean field critical behaviour for percolation in high dimensions*, Comm. Math. Phys. **128** (1990), 333–391.

[K] H. Kesten, *The critical probability of bond percolation on the square lattice equals* 1/2. Comm. Math. Phys. **74** (1980), 41–59.

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

[M] M. V. Menshikov, *Coincidence of critical points in percolation problems*, Soviet Math. Dokl. **33** (1986), 856–859.

[N] C. M. Newman, *Ising models and dependent percolation*, in Topics in Statistical Dependence (H. W. Block, A. R. Sampson, and T. H. Savits, eds.), IMS Lecture Notes–Monograph Series, to appear.

[NS] C. M. Newman and L. S. Schulman, *Infinite clusters in percolation models*, J. Statist. Phys. **26** (1981), 613–628.

[NSt] C. M. Newman and D. L. Stein, *Broken symmetry and domain structure in Ising-like systems*, Phys. Rev. Lett. **65** (1990), 460–463.

[NW] C. M. Newman and C. C. Wu, *Markov fields on branching planes*, Probab. Theory Rel. Fields **85** (1990) 539–552.

[P] R. Pemantle, *The contact process on trees*, Ann. Probab., to appear.

[PNM] M. Pokorny, C. M. Newman, and D. Meiron, *The trapping transition in dynamic (invasion) and static percolation*, J. Phys. A **23** (1990), 1431–1438.

[R] L. Russo, *On the critical percolation probabilities*, Z. Wahrsch. Verw. Gebiete **56** (1981), 229–237.

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

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

DEPT. OF MATHEMATICS, UNIVERSITY OF ARIZONA, TUCSON, AZ 85721

Current address: Courant Institute of Mathematical Sciences, New York University, New York, 10012

E-mail address: NEWMAN@CIMS19.NYU.EDU

Lectures in Applied Mathematics
Volume 27 (1991)

The Lace Expansion and the Upper Critical Dimension for Percolation

GORDON SLADE

Abstract. This talk describes some recent joint work with Takashi Hara [12], in which it is proved that the triangle condition for independent bond percolation on $\mathbf{Z}^d$ is satisfied for the nearest-neighbour model in sufficiently high dimensions ($d \geq 48$), and above six dimensions for a class of "spread-out" models. The triangle condition is known to imply that the critical exponents $\gamma, \beta, \delta, \Delta_t$ ($t = 2, 3, \ldots$) exist and take their mean-field (Bethe lattice) values. This provides further evidence that the upper critical dimension for percolation is six.

The proof uses an expansion which is related to the lace expansion for the self-avoiding walk. In this context, the lace expansion is best interpreted as arising from repeated application of the inclusion-exclusion relation.

1. The triangle condition. Consider the set of all bonds $b = \{x, y\}$, where $x, y \in \mathbf{Z}^d$ and $x \neq y$, and let $\{J_b\}$ be a fixed set of nonnegative numbers, indexed by the bonds, which is invariant under the symmetries of $\mathbf{Z}^d$. To each bond b is associated an independent Bernoulli random variable n_b, with $n_b = 1$ with probability $p \cdot J_b$ and $n_b = 0$ with probability $1 - p \cdot J_b$, where $0 \leq p \leq [\sup_b J_b]^{-1}$. The following possibilities for J_b will be considered:

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 82A43; Secondary 60K35.

Supported by NSERC grant A9351.

The detailed version of this paper will appear in Communications in Mathematical Physics.

(i) the nearest-neighbour model:

$$J_{\{x,y\}} = \begin{cases} 1 & \text{if } \|y-x\|_2 = 1\,, \\ 0 & \text{otherwise}; \end{cases}$$

(ii) the "spread-out" models:

$$J_{\{x,y\}} = L^{-d} g\left(\frac{y-x}{L}\right),$$

where L will be taken to be large, and g is a $\mathbf{Z}^d$-invariant piecewise differentiable function $g : \mathbf{R}^d \to [0, \infty)$, with $e^{\delta\|x\|_\infty} g(x) \in L^\infty(\mathbf{R}^d)$ for some $\delta > 0$. A basic example is $g(x) = I[\|x\|_\infty \le 1]$, where I is the indicator function.

If $n_b = 1$ then b is said to be *occupied*, while if $n_b = 0$ then b is said to be *vacant*. Two sites x and y are said to be *connected* (in a given configuration $\{n_b\}$) if there is a path from x to y consisting of occupied bonds. The set of sites which are connected to x is denoted $C(x)$, and the cardinality of $C(x)$ is written $|C(x)|$. Let

$$\tau_p(x, y) = \langle I[x \text{ is connected to } y]\rangle_p = \langle I[y \in C(x)]\rangle_p,$$

where $\langle\cdot\rangle_p$ denotes expectation with respect to the joint distribution of the n_b. The *triangle diagram* is defined by

$$\nabla(p) = \sum_{x,y} \tau_p(0, x)\tau_p(x, y)\tau_p(y, 0).$$

The triangle diagram can be written in terms of the Fourier transform

$$\hat{\tau}_p(k) = \sum_x \tau_p(0, x)e^{ik\cdot x} \tag{1}$$

as

$$\nabla(p) = \int_{[-\pi,\pi]^d} \frac{d^d k}{(2\pi)^d} \hat{\tau}_p(k)^3.$$

The *triangle condition* states that $\nabla(p_c) < \infty$, where $p_c \in (0, 1)$ is the critical point. The critical point has two equivalent definitions [**6**, **10**, **14**, **1**]:

$$\begin{aligned} p_c &= \sup_p\{p : \langle I[|C(0)| = \infty]\rangle_p = 0\} \\ &= \sup_p \left\{p : \chi(p) \equiv \sum_x \tau_p(0, x) < \infty\right\}. \end{aligned}$$

The triangle condition was introduced by Aizenman and Newman [**3**] as an (unverified) criterion which implies that $\chi(p) \sim (p_c - p)^{-\gamma}$

as $p \nearrow p_c$, with $\gamma = 1$, i.e., there are positive constants c_1 and c_2 such that $c_1(p_c - p)^{-1} \le \chi(p) \le c_2(p_c - p)^{-1}$, for all $p < p_c$. Subsequently Barsky and Aizenman [4] showed that if the triangle condition is satisfied then $P_\infty(p) \equiv \langle I[|C(0)| = \infty]\rangle_p \sim (p - p_c)^\beta$ as $p \searrow p_c$, with $\beta = 1$, and $M(p_c, h) \sim h^{1/\delta}$ as $h \searrow 0$ with $\delta = 2$, where $M(p, h) = 1 - \sum_{1 \le n < \infty} e^{-hn} \langle I[|C(0)| = n]\rangle_p$. Also, the triangle condition is known to imply that the gap exponents are mean field ($\Delta_t = 2$, $t = 2, 3, \ldots$) [15].

Recently it has been proved [12] that the triangle condition is satisfied in two situations as stated in the following theorem.

THEOREM 1. *The triangle condition and the infrared bound* $\hat{\tau}_p(k) \le \text{const}\cdot k^{-2}$, *uniformly in* $p < p_c$, *hold:* (a) *for the nearest-neighbour model if* d *is sufficiently large* ($d \ge 48$) *and* (b) *for* $d > 6$ *and* L *sufficiently large* (*depending on* d, g) *for the spread-out models.*

The triangle condition is expected not to hold for $d \le 6$. A consequence of the proof is that $\nu_2 = 1/2$, i.e.,

$$\left[\frac{\sum_x |x|^2 \tau_p(0, x)}{\sum_x \tau_p(0, x)}\right]^{1/2} \sim (p_c - p)^{-1/2} \quad \text{as } p \nearrow p_c.$$

Using related ideas, Hara [11] has shown that for (a) and (b) of Theorem 1, $\nu = 1/2$, i.e.,

$$-\lim_{n\to\infty} \frac{1}{n} \ln \tau_p(0, (n, 0, \ldots, 0)) \sim (p_c - p)^{1/2} \quad \text{as } p \nearrow p_c.$$

Similar methods can be used to study branched polymers above their expected upper critical dimension of eight [13].

Since the nearest-neighbor and spread-out models are expected to be in the same universality class (i.e., have the same critical exponents), Theorem 1 supports the conjecture that the triangle condition is satisfied for the nearest-neighbour model above six dimensions. Theorem 1 is complementary to the results of [9, 18], which show that all critical exponents (assuming they exist) cannot take their mean-field (Bethe lattice) values in less than six dimensions, and provides further evidence that the upper critical dimension for percolation is six. The fact that (b) holds independently of the short-range behaviour of the J_b is an illustration of universality.

Theorem 1 is proved using an expansion which is related to the Brydges–Spencer lace expansion for self-avoiding walk [8], with the convergence argument of [17]. The lace expansion is best interpreted in this context as a consequence of repeated use of the inclusion-exclusion relation. As this interpretation of the lace expansion has not been published elsewhere, it is reviewed in the next section.

2. The lace expansion and the inclusion-exclusion relation. In [**8**], the lace expansion was used to show that in more than four dimensions *weakly* self-avoiding walk behaves like simple random walk: the mean-square displacement is asymptotically linear in the number of steps. Later the lace expansion was applied directly to the *strictly* self-avoiding walk in very high dimensions [**17**]. These methods could also be used to treat "spread-out" strictly self-avoiding walk above four dimensions.

The lace expansion was first derived by an expansion and resummation procedure reminiscent of the cluster expansions of statistical mechanics and constructive quantum field theory [**7**]. Viewed differently, however, the lace expansion can be seen as resulting from repeated application of the inclusion-exclusion relation, as will now be explained. This approach to the lace expansion is similar in spirit to the work of Park [**16**] on intersection probabilities of simple random walk.

A T-step nearest-neighbour self-avoiding walk in $\mathbf{Z}^d$ is an ordered sequence $\omega = (\omega(0), \omega(1), \ldots, \omega(T))$, with $\omega(i) \in \mathbf{Z}^d$, $|\omega(i+1) - \omega(i)| = 1$, and $\omega(i) \neq \omega(j)$ for $i \neq j$. Let $\mathscr{C}_T(x, y)$ be the set of all T-step self-avoiding walks from x to y, let $c_T(x, y)$ be the cardinality of $\mathscr{C}_T(x, y)$, and let c_T be the number of T-step self-avoiding walks starting at the origin. By convention, $c_0(x, y) = \delta_{x,y}$. The two-point function is defined by

$$\sigma_z(x, y) = \sum_{T=0}^{\infty} c_T(x, y) z^T .$$

For simplicity, here z is taken to be nonnegative. Information about critical exponents, such as the one governing the mean-square displacement, is contained in the detailed behavior of the Fourier transform $\hat{\sigma}_z(k)$ (defined as in (1)) near the *critical point* z_c (the radius of convergence of the power series $\hat{\sigma}_z(0) = \sum_{T=0}^{\infty} c_T z^T$). It is this behavior which can be determined using the lace expansion, in high dimensions.

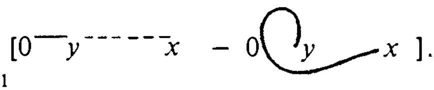
The first step in deriving the expansion is to extract the term in $\sigma_z(0, x)$ corresponding to $T = 0$,

$$\sigma_z(0, x) = \delta_{0,x} + \sum_{T=1}^{\infty} c_T(0, x) z^T . \tag{2}$$

Now for $T \geq 1$,

$$(3)\quad c_T(0, x) = \sum_{y : |y|=1} \left[c_1(0, y) c_{T-1}(y, x) - \sum_{\omega_1 \in \mathscr{C}_{T-1}(y, x)} I[0 \in \omega_1] \right].$$

Diagrammatically the right side of (3) can be represented by

$\sum_{y : |y|=1}$ 

In the first term on the right side the dotted line is unconstrained, apart from the fact that it should be self-avoiding. Equation (3) is just the inclusion-exclusion relation: the first term on the right side counts all walks from 0 to x that are self-avoiding *after* the first step, and the second subtracts the contribution due to those that are not self-avoiding from the beginning, i.e., walks that return to the origin. Since $c_1(0, y) = 1$ for $|y| = 1$, substitution of (3) into (2) gives

(4)

$$\sigma_z(0, x) = \delta_{0,x} + z \sum_{y : |y|=1} \sigma_z(y, x) - \sum_{y : |y|=1} \sum_{T=0}^{\infty} z^{T+1} \sum_{\omega_1 \in \mathscr{C}_T(y, x)} I[0 \in \omega_1].$$

The inclusion-exclusion relation can now be applied to the last term on the right side of (4), as follows. Let S be the first (and only) time that $\omega_1(S) = 0$. Then

$$\sum_{\omega_1 \in \mathscr{C}_T(y, x)} I[0 \in \omega_1] = \sum_{S=1}^{T} \sum_{\substack{\omega_2 \in \mathscr{C}_S(y, 0) \\ \omega_3 \in \mathscr{C}_{T-S}(0, x)}} I[\omega_2 \cap \omega_3 = \{0\}]$$

$$= \sum_{S=1}^{T} \left[c_S(y, 0) c_{T-S}(0, x) - \sum_{\substack{\omega_2 \in \mathscr{C}_S(y, 0) \\ \omega_3 \in \mathscr{C}_{T-S}(0, x)}} I[\omega_2 \cap \omega_3 \neq \{0\}] \right]$$

$=$ 

and hence

$$(5)\quad \sum_{y : |y|=1} \sum_{T=0}^{\infty} z^{T+1} \sum_{\omega_1 \in \mathscr{C}_T(y, x)} I[0 \in \omega_1]$$

$$= \sum_{S=2}^{\infty} z^S u_S \cdot \sigma_z(0, x) - \sum_{\substack{S=2 \\ T=0}}^{\infty} \sum_{\substack{\omega_2 \in \mathscr{U}_S \\ \omega_3 \in \mathscr{C}_T(0, x)}} z^{S+T} I[\omega_2 \cap \omega_3 \neq \{0\}]$$

where $\mathscr{U}_S$ is the set of all S-step self-avoiding loops at the origin and u_S is the cardinality of $\mathscr{U}_S$.

Continuing in this fashion, in the last term on the right side of (5) let $T_1 \geq 1$ be the first time along ω_3 that $\omega_3(T_1) \in \omega_2$, and let $v = \omega_3(T_1)$. Then the inclusion-exclusion relation can be applied again to remove the avoidance between the portions of ω_3 before and after T_1, and correct for this removal by the subtraction of a term involving a further intersection. Repetition of this procedure leads to

$$(6) \qquad \sigma_z(0, x) = \delta_{0,x} + z \sum_{y : |y|=1} \sigma_z(y, x) + \sum_v \Pi_z(0, v)\sigma_z(v, x),$$

where

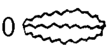
$$(7) \qquad \Pi_z(0, v) = \sum_{N=1}^{\infty} (-1)^N \Pi_z^{(N)}(0, v),$$

with

$$\Pi_z^{(1)}(0, v) = \delta_{0,v} \sum_{S=2}^{\infty} z^S u_S = \delta_{0,v} \quad 0$$


$$\Pi_z^{(2)}(0, v) = \prod_{i=1}^{3} \left[\sum_{T_i=1}^{\infty} z^{T_i} \sum_{\omega_i \in \mathscr{W}_{T_i}(0, v)} \right] I[\omega_1 \cap \omega_2 = \omega_1 \cap \omega_3$$

$$= \omega_2 \cap \omega_3 = \{0, v\}]$$

$= 0$  v.

Similarly,

$\Pi_z^{(3)}(0, v) = 0$  v,

where each wavy line represents a sum over self-avoiding walks between the endpoints of the line, weighted by z^T, with mutual avoidance between some (but not all) pairs of lines in the diagram. The unlabelled vertex is summed over $\mathbf{Z}^d$. All the higher-order terms can be expressed as diagrams in this way, and it is not difficult to see the pattern of mutual avoidance between subwalks (individual lines in the diagram) which emerges. Equation (7) is the lace expansion.

Taking the Fourier transform of (6) and solving for $\hat{\sigma}_z(k)$ gives

$$(8) \qquad \hat{\sigma}_z(k) = \frac{1}{1 - 2dz\widehat{D}(k) - \widehat{\Pi}_z(k)},$$

where

$$(9) \qquad \widehat{D}(k) = \frac{1}{d} \sum_{\mu=1}^{d} \cos k_\mu .$$

One wants now to use the lace expansion to bound $\widehat{\Pi}_z(k)$ at $z = z_c$. The first step in bounding the diagrams representing $\Pi_z^{(N)}$ is to obtain an upper bound by removing the mutual avoidance between subwalks. This relies on the *repulsive* nature of the interaction. It can then be shown [**8**] that for $N \geq 2$

$$|\widehat{\Pi}_z^{(N)}(k)| \leq [\sup_{x \neq 0} \sigma_z(0, x)] B(z)^{(N-2)/2} B_1(z)^{N/2},$$

where $B(z)$ is the "bubble diagram"

$$B(z) \equiv \sum_x \sigma_z(0, x)^2 = \int_{[-\pi, \pi]^d} \frac{d^d k}{(2\pi)^d} \hat{\sigma}_z(k)^2$$

and

$$B_1(z) = B(z) - 1 = \sum_{x \neq 0} \sigma_z(0, x)^2.$$

It is expected that $\hat{\sigma}_{z_c}(k) \leq \text{const} \cdot k^{-2}$ for k near zero (the infrared bound), and hence that $B(z_c) < \infty$ (the "bubble condition") for $d > 4$. In fact, in high dimensions it can be expected that $B_1(z_c) = O(d^{-1})$, since the main contribution should be due to $|x| = 1$ and the one-step walk to x, and this gives $z_c^2 \cdot 2d \approx (1/(2d-1))^2 \cdot 2d$. Thus one expects that $|\widehat{\Pi}_{z_c}(k)|$ will be small in high dimensions, and a more detailed analysis can be used to argue that $|\partial_{k_\mu}^2 \widehat{\Pi}_{z_c}(k)|$ should also be small. The details of how these ideas can be turned into a proof can be found in [**17**]. The analogous expansion for percolation is described in the next section, where the mechanism for using the expansion is explained in more detail.

3. Sketch of proof of Theorem 1. The proofs of parts (a) and (b) are similar, and we will discuss only the proof of part (a). Define

$$(10) \quad T(p) = \nabla(p) - 1 = \sum_{x, y} \tau_p(0, x)\tau_p(x, y)\tau_p(y, 0) - \tau_p(0, 0)^3,$$

$$(11) \quad W(p) = \sum_x |x|^2 \tau_p(0, x)^2,$$

and let T_G and W_G be the quantities obtained by replacing τ_p in (10) and (11) by the massless Gaussian propagator

$$(12) \quad C(x, y) = \sum_{\omega : x \to y} \left(\frac{1}{2d}\right)^{|\omega|} = \int_{[-\pi, \pi]^d} \frac{d^d k}{(2\pi)^d} e^{ik \cdot (y-x)} \frac{1}{1 - \widehat{D}(k)},$$

where the sum in (12) is over simple random walks, and $\widehat{D}(k)$ was defined in (9).

It follows from the continuity of τ_p in p [2] that $T(p)$ and $W(p)$ are continuous for $p < p_c$, and it can be shown that if $p \leq 1/(2d)$, then $T(p) \leq T_G \leq K_T/d$ and $W(p) \leq W_G \leq K_W/d$, with K_G and K_W independent of $d \geq 7$. (For $d \leq 6$, $T_G = W_G = \infty$.) Together with these facts, the next proposition (whose statement here is incomplete—the details are in [**12**]) implies that $T(p) \leq 3K_T/d$ for all $p < p_c$. It then follows from the continuity of τ_p in p and the monotone convergence theorem that $\nabla(p_c) \equiv 1 + T(p_c) \leq 1 + (3K_T/d)$, and hence the triangle condition is satisfied.

PROPOSITION 2 (Incomplete statement). *For the nearest-neighbour model in sufficiently high dimensions, for any fixed $p \in [1/(2d), p_c)$, if $T(p) \leq 4K_T/d$ and $W(p) \leq 4K_W/d$, then in fact $T(p) \leq 3K_T/d$ and $W(p) \leq 3K_W/d$.*

The argument which allows poor estimates on T and W to be turned into better estimates uses an expansion related to the lace expansion for the self-avoiding walk. The derivation of the expansion for percolation is more involved than for the self-avoiding walk, and we will just give the basic idea. First some definitions are needed. Given a bond configuration $\{n_b\}$ and two distinct sites x and y, a bond b is said to be *pivotal* for the connection from x to y if x and y are connected in the configuration when $n_b = 1$ but are not connected when $n_b = 0$. In the models treated in Theorem 1, at the critical point connected sites are expected typically to be connected mainly by pivotal bonds. If x and y are connected but there is no pivotal bond for the connection from x to y, then x and y are said to be *doubly connected*, denoted

$$0 \Longleftrightarrow x.$$

If 0 is connected to x then either 0 is doubly connected to x or there is a first pivotal bond (u, v) for the connection (with u doubly connected to 0). Therefore

$$(13) \quad \tau_p(0, x) = \text{Prob}(0 \Longleftrightarrow x) + \sum_{(u,v)} \text{Prob}(0 \Longleftrightarrow u\,v \longrightarrow x).$$

In the last term the connection from v to x cannot share any site in common with the cluster between 0 and u, since (u, v) is pivotal. This is a kind of repulsive interaction akin to that of the self-avoiding walk. Now one wants to proceed as for the self-avoiding walk, by extracting a factor of $\tau_p(v, x)$ from the last term on the right side of (13), and correcting with a term involving a "more connected" (and hence less important) term. The first step in this procedure is to appeal to a lemma used in [3], which implies that

$$\mathrm{Prob}(0 \Leftrightarrow uv \longrightarrow x) = p\langle I[0 \Leftrightarrow u]\tau_p^{C_{(u,v)}(0)}(v, x)\rangle_p, \tag{14}$$

where $\tau_p^{C_{(u,v)}(0)}(v, x)$ is configuration dependent and is defined as follows. The set $C_{(u,v)}(0)$ is defined to be the set of sites which remain connected to the origin after setting $n_{\{u,v\}} = 0$, and for a set A of sites, $\tau^A(v, x)$ is the probability that v is connected to x after all bonds terminating on a site in A are made vacant. Next the $\tau_p^{C_{(u,v)}(0)}(v, x)$ in (14) is replaced by

$$\tau_p(v, x) - [\tau_p(v, x) - \tau_p^{C_{(u,v)}(0)}(v, x)]. \tag{15}$$

The first term in (15) allows a factor of $\tau_p(v, x)$ to be extracted from the expectation in (14), and there is a remainder due to the second term. The treatment of the remainder is somewhat technical, and as explained in [12] leads to an iterative procedure which generates further terms in an expansion.

The result, after taking Fourier transforms and solving for $\hat{\tau}_p(k)$, is

$$\hat{\tau}_p(k) = \frac{\widehat{G}(k)}{1 - 2dp\widehat{D}(k) - \widehat{\Psi}_p(k)}, \tag{16}$$

where

$$\widehat{G}(k) = 1 + \sum_{x \neq 0} \mathrm{Prob}(0 \Leftrightarrow x)e^{ik\cdot x} + \cdots, \tag{17}$$

$$\widehat{\Psi}(k) = p \sum_{(u,v)\,:\,u\neq 0} \mathrm{Prob}(0 \Leftrightarrow u)e^{ik\cdot v} + \cdots. \tag{18}$$

The higher-order terms can be interpreted diagrammatically. By the van den Berg–Kesten inequality [5],

$$\mathrm{Prob}(0 \Leftrightarrow x) \leq \tau_p(0, x)^2,$$

and by the hypothesis in Proposition 2, $\sum_{x\neq 0} \tau_p(0, x)^2 \leq 4K_T/d$. Similar but more involved estimates requiring $T(p)$ can be made for the higher-order terms. For the self-avoiding walk the repulsive nature of the interaction was used in an important way to bound the terms of the expansion, and here that role is being played by the van den Berg–Kesten inequality. In addition, first and second derivatives with respect to k_μ of $\hat{G}$ and $\hat{\Psi}$ can be estimated using the van den Berg–Kesten inequality and the hypothesis on $W(p)$ in Proposition 2. In this way it can be shown that the hypotheses of Proposition 2 imply that $\hat{G}(k) - 1 = O(d^{-1})$ and that the denominator of (16) is a small perturbation of $(1 - \hat{D}(k))$, leading to the bound

$$\hat{\tau}_p(k) \le (1 + O(d^{-1}))(1 - \hat{D}(k))^{-1}.$$

This is the infrared bound. (Although at this point the infrared bound depends on the assumption that $T(p) \le 4K_T/d$ and similarly for $W(p)$, it will follow from Proposition 2 that in fact stronger bounds on $T(p)$ and $W(p)$ hold, and thus the infrared bound is proved.) Now by writing $T(p)$ and $W(p)$ in terms of $\hat{\tau}_p$, the improved bounds on $T(p)$ and $W(p)$ can be obtained.

Acknowledgments. I am grateful to Takashi Hara for helpful suggestions and comments on this paper, and to Michael Aizenman and to the Courant Institute for their hospitality while it was written.

REFERENCES

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

2. M. Aizenman, H. Kesten, and C. M. Newman, *Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation*, Comm. Math. Phys. **111** (1987), 505–531.

3. M. Aizenman and C. M. Newman, *Tree graph inequalities and critical behaviour in percolation models*, J. Stat. Phys. **36** (1984), 107–143.

4. D. J. Barsky and M. Aizenman, *Percolation critical exponents under the triangle condition*, preprint, 1988.

5. J. van den Berg and H. Kesten, *Inequalities with applications to percolation and reliability*, J. Appl. Prob. **22** (1985), 556–569.

6. S. R. Broadbent and J. M. Hammersley, *Percolation processes*, I. *Crystals and mazes*, Proc. Camb. Phil. Soc. **53** (1957), 629–641.

7. D. C. Brydges, *A short course on cluster expansions*, In Critical Phenomena, Random Systems, Gauge Theories (K. Osterwalder and R. Stora, eds.) Amsterdam, New York, Oxford, Tokyo, 1986, North-Holland, Les Houches, 1984.

8. D. C. Brydges and T. Spencer, *Self-avoiding walk in 5 or more dimensions*, Comm. Math. Phys. **97** (1985), 125–148.

9. J. T. Chayes and L. Chayes, *On the upper critical dimension of Bernoulli percolation*, Comm. Math. Phys. **113** (1987), 27–48.

10. J. M. Hammersley, *Bornes supérieures de la probabilité critique dans un processus de filtration*, in Le Calcul des Probabilités et ses Applications, CNRS Paris, 1959, pp. 17–37.

11. T. Hara, *Mean-field critical behaviour for correlation length for percolation in high dimensions*, Probab. Theory Rel. Fields **86** (1990), 337–385.

12. T. Hara and G. Slade, *Mean-field critical behaviour for percolation in high dimensions*, Comm. Math. Phys. **128** (1990), 333–391.

13. ——, *On the upper critical dimension of lattice trees and lattice animals*, J. Stat. Phys. **59** (1990), 1469–1510.

14. M. V. Menshikov, S. A. Molchanov, and A. F. Sidorenko, *Percolation theory and some applications*, Itogi Nauki i Tekhniki (Series of Probability Theory, Mathematical Statistics, Theoretical Cybernetics), **24** (1986), 53–110 (English Translation: J. Soviet Math. **42** (1988), 1766–1810).

15. B. G. Nguyen, *Gap exponents for percolation processes with triangle condition*, J. Stat. Phys. **49** (1987), 235–243.

16. Y. M. Park, *Direct estimates on intersection probabilities of random walks*, J. Stat. Phys. **57** (1989), 319–331.

17. G. Slade, *The diffusion of self-avoiding random walk in high dimensions*, Comm. Math. Phys. **110** (1987), 661–683.

18. H. Tasaki, *Hyperscaling inequalities for percolation*, Comm. Math. Phys. **113** (1987), 49–65.

Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada L8S 4K1

E-mail address: slade@sscvax.mcmaster.ca

Lectures in Applied Mathematics
Volume 27 (1991)

Mandelbrot Percolation in Two and Three Dimensions

GLEN H. SWINDLE

In the early 1970s Mandelbrot [M] introduced a method of generating self-similar fractals, which is now called "Mandelbrot Percolation." The scheme begins with the unit square in two dimensions, and with two numbers: N a positive integer greater than one, and p a real number in the unit interval $[0, 1]$. The unit square is divided into N^2 squares of dimension $1/N$ and each square is independently retained with probability p, or else it is removed. On the next iteration, all retained squares are again divided into N^2 subsquares, and each square is similarly retained or removed. If we denote the retained set after n iterations by A_n, then the object of interest is the set

$$A_\infty = \bigcap_{n=1}^{\infty} A_n.$$

The system was first approached rigorously by Chayes, Chayes, and Durrett. Let Ω_∞ denote the event that a crossing of the unit square exists within the limiting retained set A_∞, and define the three critical values:

$$p_\varnothing(N) = \inf\{p : P(A_\infty \neq \varnothing) > 0\},$$
$$p_d(N) = \sup\{p : P(\text{largest connected component of } A_\infty \text{is a point}) = 1\},$$
$$p_c(N) = \inf\{p : P(\Omega_\infty) > 0\}.$$

The following results, established in [CCD], identify the value of $p_\varnothing$, equate p_d with p_c (which eliminates the possibility of additional phases

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60K35.

between the "dust-like" phase and the percolation phase), and establish the discontinuity of the phase transition at p_c.

THEOREM 1. $p_{\varnothing} = 1/N^2$.

THEOREM 2. $p_d(N) = p_c(N) < 1$.

THEOREM 3. *There exists an* $\varepsilon_0 > 0$ *so that if* $P(\Omega_\infty) < \varepsilon_0$, *then* $P(\Omega_\infty) = 0$.

The result of the last theorem means that at p_c the probability of percolation across the square drops from some positive value discontinuously to zero. The driving force in this discontinuity is the asymmetry between retained squares and removed squares (once a square is removed, it stays removed forever).

The next result [CC], which we use later in the analysis of the three-dimensional system, establishes the limiting behavior of $p_c(N)$ as $N \to \infty$. In what follows, p_{site} refers to the critical probability in standard site percolation on the two-dimensional integer lattice.

THEOREM 4. $\lim_{N\to\infty} p_c(N) = p_{\text{site}}$.

We now direct our attention to the analogous three-dimensional Mandelbrot percolation system. Let N be a positive integer larger than 1, and consider the $2 \times 2 \times 1$ box $[0, 2] \times [0, 2] \times [0, 1]$ which we will denote by $\mathscr{B}_{2,2,1}$. In general, we will denote the box $[0, J] \times [0, K] \times [0, L]$ by $\mathscr{B}_{J,K,L}$. Technical considerations require that we consider the box $\mathscr{B}_{2,2,1}$ instead of the unit cube.

The procedure is to divide each of the 4 unit cubes comprising $\mathscr{B}_{2,2,1}$ into N^3 cubes of dimension $1/N$, and retain each of these cubes independently with probability p. Denote the retained set by A_1. This procedure is repeated with each of the retained cubes in A_1 to form A_2, and after n such iterations, the retained set is again denoted by A_n. As in the two-dimensional system, we are interested in properties of the limiting set

$$A_\infty = \bigcap_n A_n.$$

In [CCGS] the existence of three phase transitions is established—the first two being analogs of $p_{\varnothing}$ and p_c in the two-dimensional system, while the third is associated with the event of a crossing of the box by a "sheet." Additionally, we show that this third phase transition is also discontinuous.

We now state the results, and then outline the proofs. With $p_\varnothing$ defined as before, the first result is

THEOREM 5. $p_\varnothing = 1/N^3$.

Now let Ω_∞ denote the event that an easy crossing exists in A_∞. By an "easy" crossing we mean a crossing of $\mathscr{B}_{2,2,1}$ between the two 2×2 facets. Consider two critical values:

$$p_d = \sup\{p : P(\text{largest connected component of } A_\infty \text{ is a point}) = 1\}$$

and

$$p_c = \inf\{p : P(\Omega_\infty) > 0\}.$$

If $p \in (p_\varnothing, p_d)$ then A_∞ is "dust-like", and if $p > p_c$ then A_∞ contains strands which cross the box. The following theorem places a lower bound on p_d.

THEOREM 6. $p_d > 1/N$.

In the next result we establish that the phase transition at p_c is discontinuous and that $p_d = p_c$.

THEOREM 7. *There exists an* $\varepsilon_0 > 0$ *such that if* $P(\Omega_\infty) \leq \varepsilon_0$, *then* $P(\Omega_\infty) = 0$, *and the largest connected component is a point.*

Theorems 5 through 7 are generalizations of results established in [CCD] for the two-dimensional system. The subsequent results address issues unique to the higher-dimensional model. We define a sheet crossing of the box to be the retention of a surface crossing $\mathscr{B}_{2,2,1}$ the easy way. Topological considerations are made precise in [CCGS]; however, it suffices here to say that a surface crossing exists if the complement of the retained set A_∞^c does not contain a top to bottom crossing of $\mathscr{B}_{2,2,1}$. Let

$$p_s = \inf\{p : P(\text{exists a sheet crossing}) > 0\}.$$

We show that

THEOREM 8. $p_s(N) < 1$ *for all* N.

THEOREM 9. *For* N *sufficiently large,* $p_c(N) < p_s(N)$.

THEOREM 10. *Letting* Δ_∞ *denote the event of a sheet crossing the easy way, then there exists an* $\varepsilon_0 > 0$ *such that if* $P(\Delta_\infty) \leq \varepsilon_0$, *then* $P(\Delta_\infty) = 0$.

Theorem 8 shows that the sheet phase is indeed nontrivial. The last two results show that the strand and sheet phases are distinct (at least for large values of N), and that the phase transition at p_s is discontinuous. We now outline the proofs of Theorems 8, 9, and 10 (the proofs of Theorems 5, 6, and 7 being very similar to those in [CCD]).

Summary of the proof of Theorem 8. The idea (a modification of an argument in [CCD]) is to show that there exists an $\varepsilon > 0$ so that, if $p > 1 - \varepsilon$, then there is positive probability of a sheet crossing of $\mathscr{B}_{2,2,1}$ for all $N \geq 2$. We now restrict our attention to $N \geq 3$; the case $N = 2$ follows by comparison with $N = 4$. The approach is to consider events where after the subdivision of a cube, at least $N^3 - 1$ cubes are retained (at most one cube is lost). Loss of at most a single subcube within each cube at every iteration implies that a sheet crossing occurs.

Let $A_0 = \mathscr{B}_{1,1,1}$, and let A_m denote the retained set after m iterations. We consider the event in which at most one cube of dimension $1/N$ is lost after the first iteration:

$$G_1(A_0) = \{\text{at least } N^3 - 1 \text{ of the subcubes are retained}\}.$$

In an inductive fashion, let

$$G_2(A_0) = \{\text{at least } N^3 - 1 \text{ of the subcubes are } G_1\}$$

and

$$G_m(A_0) = \{\text{at least } N^3 - 1 \text{ of the subcubes are } G_{m-1}\}.$$

Now, let $\theta_m = P(G_m(A_0))$. It is clear that if $\theta = \lim_{m\to\infty} \theta_m > 0$, then with positive probability a sheet crossing of A_0 occurs, since removal of at most a single cube of dimension $1/N^m$ from each cube of dimension $1/N^{m-1}$ at the mth iteration never precludes a sheet crossing. The original intention was to establish a sheet crossing of $\mathscr{B}_{2,2,1}$. However, if $\theta > 0$ then the probability of a sheet crossing of $\mathscr{B}_{2,2,1}$ is greater than $\theta^4 > 0$. It is enough, therefore, to establish that for large $p < 1$, $\theta > 0$.

By construction of the events above we have

$$\theta_m = p^{N^3}\{\theta_{m-1}^{N^3} + N^3\theta_{m-1}^{N^3-1}(1-\theta_{m-1})\} + N^3 p^{N^3-1}(1-p)\theta_{m-1}^{N^3-1}.$$

The proof is completed by showing that for large values of p this equation has a positive fixed point. This is a simple computational exercise.

Summary of the proof of Theorem 9. We begin with a remark on the notion of duality in percolation. First, in two-dimensional integer lattice site percolation, duality is essentially the following: either a left-right crossing of a rectangle occurs by occupied sites, or else there is a top-bottom crossing (with a modified notion of connectedness) of the rectangle by vacant sites (see [D] for details). In the three-dimensional system we need the following two dual statements (see [CCGS]).

(a) Either a left-right crossing of a box occurs in A_∞, or else there is a top-bottom sheet crossing of the box in A_∞^c (defined with an appropriate notion of connectedness) which separates the left face from the right face of the box.

(b) Either a top-bottom crossing of the box occurs in A_∞^c (again, with appropriately defined connectedness) or else a sheet crossing occurs in A_∞ which separates the top face from the bottom face of the box.

We prove Theorem 9 by showing that for large values of N there are values of p so that $P_p(\Omega_\infty) > 0$ (there is positive probability of percolation in the retained set), at the same time that a dual crossing (which precludes sheet percolation) occurs almost surely. With this goal in mind, after n Mandelbrot iterations we will focus our attention on a single layer of cubes L_n—those cubes of dimension $1/N^n$ adjacent to the plane $y = 0$. Note that, after n iterations, the probability of a dual crossing entirely within this layer is greater than the probability that the dual of integer lattice site percolation with parameter p crosses a box of size $N^n \times 2N^n$ the hard way, since conditioning on complete retention reduces the probability of a dual crossing.

The next observation to make is that there is an effective enhancement of this dual crossing event due to the layer of cubes adjacent to L_n. Specifically, if five cubes in the next layer arranged as in Figure 1 are removed (this occurs with probability $(1-p)^5$), then the site in L_n adjacent to the central cube in this arrangement is effectively removed, in the sense that retention of this cube does not prevent a dual crossing. This enhancement and the convergence of $p_c(N)$ to p_{site} are the key ingredients to our proof. The idea is to pick N large enough so that $p_c(N)$ is close enough to p_{site} so that the enhancement of the dual crossings pushes the dual crossing probability to one.

To make this precise, we consider the sublattice $V = 3Z$, in which sites are separated by a minimum distance of three units. Sites in V will be those sites which have lowered occupation probability (enhanced dual occupation), and these enhancing events are *independent*.

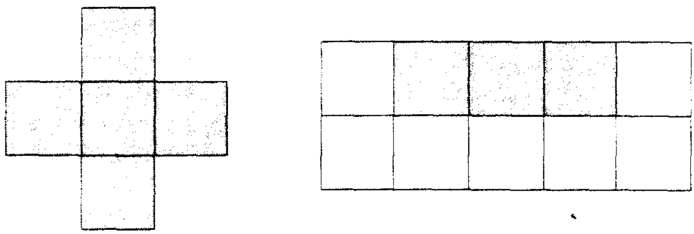


FIGURE 1. Enhancement of dual crossing

In Chapter 10 of [**K**], conditions are given under which site percolation is forced into the subcritical regime by lowering the occupation probabilities on a periodic sublattice of a periodic two-dimensional graph. The result that we are going to use is Theorem 10.2 of [**K**], which, in this particular scenario, can be stated as

LEMMA 9.1. *Let* $P_{p_{\text{site}}}$ *denote the measure on occupation configurations in two-dimensional site percolation. Let* $P_{p'}$ *denote a similar measure with site occupation probabilities equal to* p_{site} *except on the sublattice* V, *where they are strictly less than* p_{site}. *Then,*

$$E_{p'}\{|C_0|\} < \infty \quad \text{and} \quad P_{p'}\{|C_0| = \infty\} = 0,$$

where $|C_0|$ *denotes the number of sites connected to the origin.*

Lemma 9.1 follows from Theorem 10.2 of [**K**] upon verification of Condition D in Chapter 10 of [**K**], which is satisfied in this case. We will use one additional result which appears as Corollary 5.1 of [**K**].

LEMMA 9.2. $\{p : 0 < p \text{ and } E_p\{|C_0|\} < \infty\}$ *is open.*

Theorem 9 now follows easily from the preceding lemmas. Lemmas 9.1 and 9.2 imply that there is a $\delta > 0$, so that if $|p - p_{\text{site}}| < \delta$, then the "enhanced" site percolation problem is subcritical. This implies that with probability one, after a large number of iterations there will exist a top-bottom dual crossing of the box $[0, 1] \times \{0\} \times [0, 2]$ (really $[0, N^n] \times \{0\} \times [0, 2N^n]$ after n iterations). Now, by Theorem 4, N can be selected large enough so that $|p_c(N) - p_{\text{site}}| < \delta/2$, which implies that there are values of p so that there is positive probability of crossing in A_∞, but with probability one that a dual crossing exists in A_∞^c, preventing a sheet crossing.

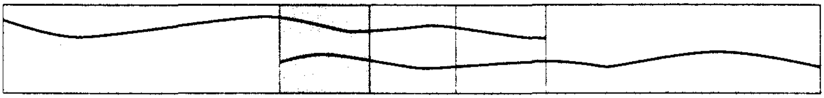


FIGURE 2. Renormalization of dual crossing events

Summary of the proof of Theorem 10. Denote the event of a dual crossing of the box $\mathscr{B}_{k,l,m}$ between the $k \times l$ faces by $\mathscr{C}_{k,l,m}$. The result is a consequence of two lemmas. The first lemma, loosely speaking, states that if the probability of a dual crossing of $\mathscr{B}_{1,2,2}$ is highly likely, then so is the probability of a dual crossing of $\mathscr{B}_{1,1,M}$. This result is used with the second lemma to show that if the probability of a dual crossing of $\mathscr{B}_{1,2,2}$ is close enough to one, then it is in fact one.

LEMMA 10.1. *If* $P(\mathscr{C}_{1,2,2}) \geq 1 - \varepsilon$, *then* $P(\mathscr{C}_{1,1,M}) \geq 1 - f_M(\varepsilon)$ *where* f_M *is a function such that* $f_M(\varepsilon) \to 0$ *as* $\varepsilon \to 0$.

LEMMA 10.2. *There exists an integer* M *and an* ε_0, *so that if* $P(\mathscr{C}_{1,1,2M}) \geq 1 - \varepsilon_0$, *then* $\lim_{n\to\infty} P(\mathscr{C}_{1,1,2^n M}) = 1$.

REMARK ON LEMMA 10.1. The second lemma is used to show that once a dual crossing of a long box ($\mathscr{C}_{1,1,M}$) is highly likely, then the dual crossing of a longer box is, in fact, more likely. The first lemma is what tells us that $\mathscr{C}_{1,1,M}$ is likely if $\mathscr{C}_{1,1,2}$ occurs with high probability. The proof is an application of Harris' inequality.

REMARK ON LEMMA 10.2. The idea is to place $4N-1$ of the boxes $\mathscr{B}_{1,1,2M}$ colinearly with each box overlapping its neighbors on an interval of length M (see Figure 2, which shows two $\mathscr{B}_{1,1,6}$ boxes with $M = 3$ overlap). Then with probability exceeding $(1-\varepsilon_0)^{4N-1}$ all of these boxes have dual crossings the long way due to the fact that these events are increasing. Consider any two of the overlapping translates of $\mathscr{B}_{1,1,2M}$. The overlapping region contains M cubes of unit size, and the probability that at least one of these cubes is removed is at least $(1-p^M)$. Removal of one of these boxes connects the two original dual crossings. Consequently,

$$P(\mathscr{C}_{1,1,2MN}) \geq (1-\varepsilon)^{4N-1}(1-p^M)^{4N-2}.$$

Now consider $N^2 \mathscr{B}_{1,1,4MN}$ boxes arranged to form a $\mathscr{B}_{N,N,4MN}$. Clearly a crossing *within* any one of these N^2 component boxes implies a crossing of the larger box, and these events are independent.

Therefore, after rescaling the resulting box $\mathscr{B}_{N,N,4MN}$, we see that

$$P(\mathscr{C}_{1,1,4M}) \geq 1 - \{1 - (1-\varepsilon)^{(4N-1)}(1-p^M)^{(4N-2)}\}^{N^2}$$
$$= 1 - \varepsilon'$$

where

$$\varepsilon' \leq \{(4N-1)\varepsilon + (4N-2)p^M\}^{N^2}$$
$$\leq (4N)^{N^2}\{\varepsilon + p^M\}^{N^2}.$$

We now iterate this procedure, with component boxes $\mathscr{B}_{1,1,4M}$, and at the nth iteration with boxes $\mathscr{B}_{1,1,2^n M}$. The overlap of the component boxes is 2^{n-1}. Note that this increases in n, and, therefore, the probability that no overlap box is removed decreases exponentially to zero. In an identical fashion, we find

$$P(\mathscr{C}_{1,1,2^{n+1}M}) \geq 1 - \varepsilon_{n+1}$$

where

$$\varepsilon_{n+1} \leq (4N)^{N^2}\{\varepsilon_n + (p^M)^{2^{n-1}}\}^{N^2}.$$

Consequently, if ε_0 is sufficiently small, and M is sufficiently large, $\lim_{n\to\infty}\varepsilon_n = 0$, completing the proof.

References

[CC] J. T. Chayes and L. Chayes, *The large N limit of the threshold values in Mandelbrot's percolation process*, J. Phys. A: Math. Gen. **22** (1989), L501–L506.

[CCGS] J. T. Chayes, L. Chayes, E. Grannan, and G. Swindle, *Critical properties of Mandelbrot's percolation process in three dimensions* (1989), submitted.

[CCD] J. T. Chayes, L. Chayes, and R. Durrett, *Connectivity properties of Mandelbrot's percolation process*, Probab. Theory Related Fields (1988).

[D] R. Durrett, *Lecture Notes on particle systems and percolation*, Wadsworth, Inc., Belmont, CA, 1988.

[K] H. Kesten, *Percolation theory for mathematicians*, Birkhauser, 1982.

[M] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Co., New York, 1983.

University of California, Los Angeles, CA
E-mail address: swindle @ bernoulli.ucsb.edu

Lectures in Applied Mathematics
Volume 27 (1991)

Intersection Probabilities for Random Walks

GREGORY F. LAWLER

1. Introduction

Let A be a finite connected subset of Z^d $(d \geq 2)$ containing 0 and let S_t be a simple random walk taking values in Z^d. There are two closely related measures of the behavior of the set A at the point 0:

- the harmonic measure of 0, i.e., the probability that a random walker "starting at ∞" conditioned to hit A first hits A at the point 0;
- the probability that $S_t \notin A$ for $t = 1, \ldots, n^2$ where $S_0 = 0$ and $n = [\mathrm{rad}(A)] = \sup\{|x| : x \in A\}$.

A number of problems have arisen which are equivalent to considering one of the above quantities for some set A. In some cases the set A is random and changing with time. We list three such problems here.

Diffusion limited aggregation. In this model for dendritic growth first introduced by Witten and Sander [**16**], we have a Markov chain whose state space is the set of finite connected subsets of Z^d containing 0. We start with $A_0 = \{0\}$ and the transition probability is

$$P\{A_{n+1} = A_n \cup \{x\} \mid A_n\} = \begin{cases} 0 & x \notin \partial A_n \\ H_{\partial A_n}(x) & x \in \partial A_n. \end{cases}$$

Here $\partial A_n = \{x \in Z^d \backslash A_n : |x - y| = 1$ for some $y \in A_n\}$ and $H_{\partial A_n}(x)$ denotes the harmonic measure. Many simulations of diffusion limited

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 60J15.

Research supported by NSF grant DMS-8702620 and an Alfred P. Sloan Research Fellowship.

aggregation (DLA) clusters have been made. For $d = 2$ it is conjectured that the "fractal dimension" of the cluster is about 1.7; in particular, the clusters are rather sparse and do not eventually fill up Z^2. Kesten [7], [8] proved that almost surely for n sufficiently large,

$$\text{rad}(A_n) \leq \begin{cases} cn^{2/3} & d = 2 \\ cn^{2/d} & d \geq 3. \end{cases} \tag{1}$$

Random walk intersections. Let A be the random set $S^1[0, n]$, where $S^1_\cdot$ is a random walk starting at 0 independent of $S_\cdot$. The problem is to determine the probability that the paths of the two random walks intersect. If we let $f(n) = P\{S^1[0, n] \cap S(0, n] \neq \emptyset\}$, then [10], [4], [2], [3]

$$f(n) \approx \begin{cases} c > 0 & d \geq 5 \\ (\log n)^{-1/2} & d = 4 \\ n^{-\zeta_d} & d = 2, 3. \end{cases}$$

(Here $\approx$ means that the logarithms of both sides are asymptotic.) For $d = 2, 3$, the best rigorous estimates of the exponents are:

$$\frac{1}{2} + \frac{1}{8\pi} \leq \zeta_2 < \frac{3}{4}, \tag{2}$$

$$\frac{1}{4} \leq \zeta_3 < \frac{1}{2}. \tag{3}$$

Duplantier and Kwon [5] have conjectured from a conformal invariance argument that $\zeta_2 = 5/8$; they also have Monte Carlo simulations which are consistent with this conjecture. Monte Carlo simulations by Burdzy, Polaski, and the author [4] suggest $\zeta_2 \approx 0.62$ (a little lower than the conjecture—however, the rate of convergence seems to be slow so we cannot reject the conjecture of Duplantier and Kwon) and $\zeta_3 \approx 0.29$.

There are a number of reasons to be interested in finding ζ_2 and ζ_3. This is one of the easiest stated "critical exponents" which has many of the characteristics of exponents from mathematical physics: (a) it is dimension dependent; (b) it takes on nontrivial values below the critical dimension $d = 4$; and (c) it is possible that it takes on a rational value for $d = 2$ but an irrational value for $d = 3$. This problem could be a good test for some nonrigorous techniques used in mathematical physics, e.g., renormalization and conformal invariance. This exponent is also the simple random walk analogue to an exponent for self-avoiding walks, see [11].

Loop-erased or Laplacian self-avoiding random walk. This is a model for self-avoiding walk (SAW) first introduced by the author [**9**] and later independently by Lyklema and Evertsz [**6**]. There are two equivalent definitions—we give here the one similar to DLA. We consider the Markov chain whose state space is the set of SAWs $\omega = [\omega_0, \dots, \omega_n]$ starting at the origin in Z^d. Let

$$\omega^0 = [0],$$

and let the transition probability be given as follows for $d \geq 3$ (the transition is slightly different for $d = 2$). If $|x - \omega_n| = 1$,

$$P\{\omega^{n+1} = [\omega_1, \dots, \omega_n, x] \mid \omega^n = \omega\} = \frac{\mathrm{Es}(x, \{\omega_0, \dots, \omega_n\})}{\sum_{|y-\omega_n|=1} \mathrm{Es}(y, \{\omega_0, \dots, \omega_n\})},$$

where

$$\mathrm{Es}(x, A) = P^x\{S_t \notin A, \ t = 0, 1, 2, \dots\}.$$

The interesting exponent $\alpha = \alpha_d$ for this model (for $d = 2, 3$) is defined by

$$\langle|\omega^n|^2\rangle \approx n^{\alpha_d}.$$

It has been proved that α is at least as large as the Flory exponents for the usual self-avoiding walk model, i.e., $\alpha_d \geq \frac{6}{2+d}$, but the exponent is expected to be larger than this value.

In this paper I will review some of the rigorous work which has been done recently in the area of intersections of random walks. Sections 2 and 3 contain some basic ideas that are useful in this subject. We first rigorously state the relationship between harmonic measure and escape probabilities in Section 2. A useful technical tool is the discrete Harnack's inequality. In Section 3 we discuss the case where A_n is a random set with a translation invariant measure. In the case of random walk intersections, it is easier to consider intersections of a one-sided walk and a two-sided walk than two one-sided walks because the measure on two-sided walks is translation invariant. The last section summarizes recent results of Burdzy and myself in establishing the inequalities (2) and (3); in particular, the exponent ζ_d is related to the rate of exponential decay in a large deviations problem. I will not discuss either DLA or the Laplacian SAW here. Readers interested in those results should see [7], [**8**], and [**12**] respectively.

2. Escape probabilities and harmonic measure. Let A be a finite, connected subset of Z^d $(d \geq 2)$, and $\text{rad}(A) = \sup\{|x| : x \in A\}$. If $S_.$ denotes a simple random walk in Z^d, we define the hitting time of A by

$$\tau_A = \inf\{t \geq 1 : S_t \in A\}.$$

For $(x, y) \in Z^d \times A$ we let

$$\overline{H}_A(x, y) = P^x\{\tau_A < \infty, S_{\tau_A} = y\},$$
$$H_A(x, y) = \overline{H}_A(x, y)[P^x\{\tau_A < \infty\}]^{-1}.$$

If $d = 2$, $P^x\{\tau_A < \infty\} = 1$ and $\overline{H}_A(x, y) = H_A(x, y)$.

The (discrete) harmonic measure of a set A is defined by

$$H_A(y) = \lim_{|x|\to\infty} H_A(x, y).$$

The limit is well known to exist [15]. There is a relationship between the harmonic measure of a point $y \in A$ and the probability that a random walk starting at y avoids A. There are at least four natural ways to measure the harmonic measure of escape probability of a point $y \in A$, if $\text{rad}(A) = n$. Before listing them we need two more definitions. Let λ_n be the hitting time of the sphere of radius n, i.e., $\lambda_n = \inf\{t : |S_t| \geq n\}$. For each $0 < r < 1$, let T_r be a geometric random variable independent of $S_.$ with rate r, i.e., $P\{T_r = j\} = (1 - r)^j r$. We think of T_r as being the killing time of the random walk with killing rate r. Then the four quantities of interest are:

$$\text{(I)} \quad P^y\{\tau_A > n^2\},$$
$$\text{(II)} \quad P^y\{\tau_A > T_{1/n^2}\},$$
$$\text{(III)} \quad P^y\{\tau_A > \lambda_{2n}\},$$
$$\text{(IV)} \quad H_A(y).$$

What we would like to show is that all four are in some sense equivalent. For (I) and (II) we state without proof an easy Tauberian theorem.

PROPOSITION 2.1. *Suppose a_n is a decreasing sequence of nonnegative numbers and $\alpha \in (0, 1)$. Let $R(r) = \sum_{j=0}^{\infty}(1 - r)^j r a_j$.*

(a) *Suppose $c_1 n^{-\alpha} \leq a_n \leq c_2 n^{-\alpha}$ for some $0 < c_1 < c_2 < \infty$. Then there exist $k_1(c_1, c_2, \alpha)$ and $k_2(c_1, c_2, \alpha)$ such that*

$$k_1 n^{-\alpha} \leq R\left(\frac{1}{n}\right) \leq k_2 n^{-\alpha}.$$

(b) *Suppose* $k_1 n^{-\alpha} \le R(\frac{1}{n}) \le k_2 n^{-\alpha}$ *for some* $0 < k_1 < k_2 < \infty$. *Then there exist* $c_1(k_1, k_2, \alpha)$ *and* $c_2(k_1, k_2, \alpha)$ *such that*

$$c_1 n^{-\alpha} \le a_n \le c_2 n^{-\alpha}.$$

We note that the assumption $\alpha < 1$ is needed for the proposition. For example, if $a_0 = 1$ and $a_n = 0$ for $n \ge 1$, then $R(1/n) = 1/n$. The proposition can be applied immediately to (I) and (II) by setting $a_n = P^y\{\tau_A > n\}$. Then $R(r) = P^y\{\tau_A > T_r\}$. We also note that an analogous proposition would hold if we assumed

$$c_1 n^{-\alpha} F(n) \le a_n \le c_2 n^{-\alpha} F(n),$$

where F is some slowly varying function of n.

We now consider (I) and (III). Let

$$c = \inf_{m>1} P^0 \left\{ |S_j| \le \frac{m}{2}, j = 0, 1, \dots, m^2 \right\} > 0.$$

Then the Markov property gives $P^y\{\tau_A > n^2 \mid \tau_A > \lambda_{2n}\} \ge c$ and hence

$$P^y\{\tau_A > n^2\} \ge c P^y\{\tau_A > \lambda_{2n}\}. \tag{4}$$

There is no general inequality in the opposite direction; in fact, it is easy to construct connected sets A with $\mathrm{rad}(A) = n$ such that $P^y\{\tau_A > n^2\} > 0$ and $P^y\{\tau_A > \lambda_{2n}\} = 0$. However, there is an inequality in the other direction that is useful.

PROPOSITION 2.2. *For every* $0 < c_1 < c_2 < \infty$, $\alpha > 0$, *there exists* $k(c_1, c_2, \alpha)$ *such that if* $\mathrm{rad}(A) \le n$, $y \in A$, $P^y\{\tau_A > n^2\} \le c_2 n^{-\alpha}$, *and for all* $m \ge n^2$, $P^y\{\tau_A > m^2\} \ge c_1 m^{-\alpha}$, *then*

$$P^y\{\tau_A > \lambda_{2n}\} \ge k n^{-\alpha}.$$

PROOF. We may assume $n \ge 5$. Let J be a positive integer and let $\rho < 1$ be defined by

$$\rho = \sup_{5 \le m < \infty} \sup_{|x| \le 2m} P^x\{\lambda_{2m} \ge m^2\} \le \sup_{5 \le m < \infty} P^0\{\lambda_{4m} > m^2\}.$$

Then by the Markov property,

$$P^y\{\tau_A > Jn^2, \lambda_{2n} > Jn^2\} \le P^y\{\tau_A \ge n^2\}\rho^{J-1} \le c_2 \rho^{J-1} n^{-\alpha}.$$

Therefore,

$$\begin{aligned} P^y\{\tau_A > \lambda_{2n}\} &\geq P^y\{\lambda_{2n} \leq Jn^2, \tau_A > Jn^2\} \\ &= P^y\{\tau_A > Jn^2\} - P^y\{\lambda_{2n} > Jn^2, \tau_A \geq Jn^2\} \\ &\geq c_1(\sqrt{J}n)^{-\alpha} - c_2\rho^{J-1}n^{-\alpha}. \end{aligned}$$

Choose J so that $k = c_1(\sqrt{J})^{-\alpha} - c_2\rho^{J-1} > 0$, and the proposition is proved. □

There is a nice inequality which relates (III) and (IV). We start by stating a very useful inequality for working with hitting probabilities.

THEOREM 2.3 (Harnack's Inequality). *For every* $1 < r < R < \infty$, *there exist* $c_1(r, R) < \infty$ *and* $c_2(r) < \infty$, *such that if* $\text{rad}(A) \leq n$, $y \in A$,

(a) $c_1^{-1}\overline{H}^A(x, y) \leq \overline{H}^A(z, y) \leq c_1\overline{H}^A(x, y)$, $rn \leq |x|, |z| \leq Rn$.
(b) $c_2^{-1}H^A(x, y) \leq H^A(z, y) \leq c_2H^A(x, y)$, $rn \leq |x|, |z| \leq \infty$.

A proof of this inequality can be found in [14]. It is also shown in that paper that one can prove the existence of harmonic measure as a consequence of this result. Another corollary of the inequality is the following.

PROPOSITION 2.4. *There exist* c_1, c_2 *such that if* $\text{rad}(A) \leq n$, $y \in A$, *then*

$$c_1H(y) \leq P^y\{\lambda_{2n} < \tau_A\}[P^z\{\lambda_{2n} > \tau_A\}]^{-1}n^{2-d} \leq c_2H(y),$$

where $z = z_n = ([\frac{3}{2}n], 0, \ldots, 0)$.

The proof can be found in [4]. If $d = 2$ and A is a connected set of radius n containing 0, one can show (see e.g. [4]) that

$$0 < c_1 < P^z\{\lambda_{2n} < \tau_A\} < c_2 < 1,$$

where c_1, c_2 are constants independent of n and A. Therefore

$$c_1H(y) \leq P^y\{\lambda_{2n} < \tau_A\} \leq c_2H(y)$$

for appropriately chosen c_1, c_2.

3. Translation invariant sets. In this section it will be convenient to consider criterion (II) or in general to consider $P^y\{\tau_A > T_r\}$. The easiest case is when $A = \{0\}$. If we let $\sigma = \sup\{t \leq T_r : S_t = 0\}$, then

$$(5)\qquad \begin{aligned} 1 &= \sum_{t=0}^{\infty} P^0\{\sigma = t\} \\ &= \sum_{t=0}^{\infty} P^0\{S_t = 0,\, t \leq T_r\} P^0\{S_s \neq 0 \text{ for } 1 \leq s \leq T_r\} \\ &= P^0\{\tau_A > T_r\} g_r(0), \end{aligned}$$

where g_r is the standard Green's function

$$g_r(x) = \sum_{t=0}^{\infty} P^0\{S_t = x,\, t \leq T_r\}.$$

It is standard that as $n \to \infty$, $g_{1/n}(0) \sim cn^{1/2}$, $d = 1$; $g_{1/n}(0) \sim c(\log n)$, $d = 2$; $g_{1/n}(0) \sim c$, $d > 2$. Therefore, we get the behavior of $P^0\{\tau_A > T_r\}$. The same idea can easily be applied to $A = \{(x, 0, \dots, 0) : x \in Z\} \subset Z^d$. Singleton sets and lines have the property that they "look the same" at each point. These ideas can be extended to the case where A is a stochastically translation invariant set.

Suppose $[x]_n = \{\dots, x_{-1}, x_0, x_1, \dots\}$ is a two-sided sequence of points in $Z^d \cup \{\infty\}$. We define $[T_k x]$ and $[-x]$ to be the sequences

$$\begin{aligned} [T_k x]_n &= x_{n+k} - x_k, \\ [-x]_n &= x_{-n}, \end{aligned}$$

where $\infty + \infty = \infty - \infty = \infty + x = \infty - x = \infty$. Note that if $[x]_0 = 0$, then $[T_k x]_0 = 0$ if $x_k \neq \infty$. Let $A(n)$, $n \in Z$, be a stochastic process taking values in Z^d satisfying

(a) $A(0) = 0$.
(b) If $n > 0$ and $A(n) = \infty$, then $A(m) = \infty$ for $m \geq n$; if $n < 0$ and $A(n) = \infty$, then $A(m) = \infty$ for $m \leq n$.
(c) If $[x]_n$ is a sequence with $x_k \neq \infty$, then

$$\begin{aligned} &P\{A(n) = [x]_n,\, m_1 \leq n \leq m_2\} \\ &\qquad = P\{A(n) = [T_k x]_n,\, m_1 - k \leq n \leq m_2 - k\}; \end{aligned}$$

(d) If $[x]_n$ is any sequence,

$$P\{A(n) = [x]_n,\, m_1 \leq n \leq m_2\} = P\{A(n) = [-x]_n,\, -m_2 \leq n \leq -m_1\}.$$

Examples of such sets are as follows.

(a) A straight line: with probability $1/2$ let $A(n) = (n, 0, \ldots, 0)$ for each n, and with probability $1/2$ let $A(n) = (-n, 0, \ldots, 0)$ for each n.

(b) Let P be any probability measure on the set of finite subsets of Z^d containing 0 which satisfies

$$P(A) = P(A - x), \quad \text{whenever } x \in A.$$

(Here $A - x = \{y - x : y \in A\}$.) Any such P can easily be put into the form of an $A(n)$ as above. An example is the set of lattice animals, i.e., the set of all finite subsets of Z^d with measure

$$P_\beta(A) = Z^{-1}\beta^{|A|},$$

where $|\cdot|$ denotes cardinality. If c_n is the number of lattice animals of cardinality n, and $\overline{\beta} = \lim_{n\to\infty}(c_n)^{1/n}$, then P_β is a probability measure for $\beta < \overline{\beta}$ with $Z = \sum_{n=1}^{\infty} c_n \beta^n$.

(c) Suppose $S^1_\cdot$ and $S^2_\cdot$ are independent simple random walks starting at 0 with killing rate r, and corresponding killing times T^1_r, T^2_r. Let A be the corresponding "two-sided" walk

$$A(n) = \begin{cases} S^1_n & 0 \le n \le T^1_r, \\ S^2_{-n} & -T^2_r \le n \le 0, \\ \infty & \text{otherwise.} \end{cases}$$

Now let $S_\cdot$ be a random walk, independent of A, with killing rate r and killing time T_r, and let g_r be the Green's function for the walk. For any set $B \subset Z^d$, let $\mathrm{Es}(B) = P^0\{\tau_B > T_r\}$. If $B \subset Z^d \cup \{\infty\}$, let $\mathrm{Es}(B) = \mathrm{Es}(B \cap Z^d)$. Define the following random variables (which depend on r):

$$J_m = \mathrm{Es}[\{A(n) : n \ge m\}],$$

$$J = \mathrm{Es}[\{A(n) : n \in Z\}],$$

$$G = \sum_{n=-\infty}^{\infty} g_r(A(n)), \quad \text{where } g_r(\infty) = 0,$$

D = indicator function of the event

$$\{A(n) \neq 0 : n = 1, 2, \ldots\}.$$

We can then prove two identities [**13**]:

PROPOSITION 3.1. *If A is transient, i.e.,*

$$P\{\exists x \in Z^d \text{ with } A(n) = x \text{ i.o.}\} = 0,$$

then

$$E(DJ) = E(DJ_0J_1).$$

PROPOSITION 3.2. *If $E(G) < \infty$, then*

$$E(JGD) = 1.$$

Note that (5) is a special case of Proposition 3.2, where $A(0) = 0$, $A(n) = \infty$, $n \neq 0$. We now give an example using each of these propositions. Suppose $A(n)$ is a line as in Example (a). Then $D = 1$, and J is not random. Let $A = \{(n, 0, \dots, 0) : n \in Z\}$ and $A^m = \{(n, 0, \dots, 0); n \geq m\}$. Then Proposition 3.1 gives

$$P^0\{\tau_A > T_r\} = P^0\{\tau_{A^0} > T_r\}P^0\{\tau_{A^1} > T_r\}.$$

As mentioned before it is easy to show that as $n \to \infty$,

$$P^0\{\tau_A > T_{1/n}\} \sim \begin{cases} cn^{-1/2} & d = 2, \\ c(\log n)^{-1} & d = 3, \\ c & d > 3. \end{cases}$$

It is also not difficult to show [13] that

$$P^0\{\tau_{A^0} > T_{1/n}\} \geq cP^0\{\tau_{A^1} > T_r\},$$

where c is independent of r. We therefore get

$$c_1 n^{-1/4} \leq P^0\{\tau_{A_0} > T_{1/n}\} \leq c_2 n^{-1/4}, \quad d = 2,$$
$$c_1(\log n)^{-1/2} \leq P^0\{\tau_{A_0} > T_{1/n}\} \leq c_2(\log n)^{-1/2}, \quad d = 3.$$

The estimate for $d = 2$ (actually a very similar estimate) was first derived by Kesten [8] using a different argument. In fact, he proved a stronger result: there exists a constant $c > 0$ such that if A is any finite connected subset of Z^2 containing 0 with $\text{rad}(A) \geq n^{-1/2}$, then

$$P^0\{\tau_A > T_{1/n}\} \leq cn^{-1/4}.$$

We now consider the case where A is a two-sided random walk with killing rate $r = r_n = 1/n$, and the walk $S.$ also has killing rate r_n. Then

$$E(J) = P\{(S^1[0, T^1_{r_n}] \cup S^2[0, T^2_{r_n}]) \cap S(0, T_{r_n}] = \emptyset\},$$

i.e., the probability that a two-sided walk and a one-sided walk with killing rate $1/n$ do not intersect. Proposition 3.2 gives

$$E(JGD) = 1. \tag{6}$$

It is relatively easy to show that as $n \to \infty$,

$$E(G) \sim \begin{cases} cn^{\frac{4-d}{2}} & d < 4, \\ c(\log n) & d = 4, \\ c & d > 4. \end{cases}$$

Also $E(D) \sim c(\log n)^{-1}$ for $d = 2$, and $E(D) \sim c > 0$ for $d > 2$. It is tempting to use (6) to conclude that

$$E(J) \approx [E(G)]^{-1}[E(D)]^{-1};$$

in fact, one can get such a result and prove

$$c_1 n^{(d-4)/2} \le E(J) \le c_2 n^{(d-4)/2}, \quad d = 2, 3, \tag{7}$$

$$c_1(\log n)^{-1} \le E(J) \le c_2(\log n)^{-1}, \quad d = 4. \tag{8}$$

(It is straightforward to show that $E(J) > c > 0$ for $d > 4$.) Note that the logarithmic term $E(D)$ has disappeared for $d = 2$. This happens because a random walk in Z^2, conditioned so that another random walk never hits its path, does not return to 0 very often.

4. Intersections of random walks. We now return to the problem of intersections of random walks starting at the origin, i.e.,

$$f(n) = P\{S^1(0, n] \cap S^2[0, n] = \varnothing\}.$$

From (7) and (8), or the equivalent result for walks of fixed length, if

$$F(n) = P\{S^1(0, n] \cup (S^2[0, n] \cap S^3[0, n]) = \varnothing\},$$

then

$$\begin{aligned} c_1 n^{(d-4)/2} &\le F(n) \le c_2 n^{(d-4)/2}, \quad d < 4, \\ c_1(\log n)^{-1} &\le F(n) \le c_2(\log n)^{-1}, \quad d = 4. \end{aligned} \tag{9}$$

A simple application of the Schwarz inequality gives $F(n) \le f(n) \le \sqrt{F(n)}$; so

$$\begin{aligned} c_1 n^{(d-4)/2} &\le f(n) \le c_2 n^{(d-4)/4}, \quad d = 2, 3, \\ c_1(\log n)^{-1} &\le f(n) \le c_2(\log n)^{-1/2}, \quad d = 4. \end{aligned} \tag{10}$$

For $d = 4$, the right-hand inequality is almost sharp [10] in the sense that

$$\lim_{n\to\infty} \frac{\log f(n)}{\log\log n} = -\frac{1}{2}. \tag{11}$$

In fact, it is probably true that $f(n) \sim c(\log n)^{-1/2}$ but this has not been proven. The proof of (11) is technical but makes use of the fact that "short range" intersections and "long range" intersections of four-dimensional paths are almost independent. We do not expect this behavior in less than four dimensions and hence would not expect the inequalities to be sharp.

For the remainder of this section we will consider only dimensions two and three. Let $S^1_\cdot, S^2_\cdot, \ldots$ be independent simple random walks starting at 0 and let

$$f(n, j) = f_d(n, j) = P\{S^1(0, n] \cap (S^2[0, n] \cup \cdots \cup S^{j+1}[0, n]) = \varnothing\}.$$

Then $f(n) = f(n, 1)$ and $F(n) = f(n, 2)$. The intersection exponents are defined by

$$\zeta(j) = \zeta_d(j) = \lim_{n\to\infty} -\frac{\log f(n, j)}{\log n},$$

i.e., $f(n, j) \approx n^{-\zeta(j)}$. It has been proved [2] that the exponents exist and are equal to corresponding exponents for Brownian motion: let $X^1_\cdot, X^2_\cdot, \ldots$ be independent Brownian motions starting at 0 in R^d, and let $\tau^i_a = \inf\{t > 0 : |X^i_t| = a\}$. Let

$$p(\varepsilon, j) = P\{X^1[\tau^1_\varepsilon, \tau^1_1] \cap (X^2[\tau^2_\varepsilon, \tau^2_1] \cup \cdots \cup X^{j+1}[\tau^{j+1}_\varepsilon, \tau^{j+1}_1]) = \varnothing\}$$

and

$$\xi(j) = \xi_d(j) = \lim_{\varepsilon\to 0} \frac{\log p(\varepsilon, j)}{\log \varepsilon}.$$

Then $\xi(j) = 2\zeta(j)$. It will be easier to consider $\xi(j)$ at this point. From (9) we get $\xi_2(2) = 2\zeta_2(2) = 2$, $\xi_3(2) = 2\zeta_3(2) = 1$.

For any $\varepsilon > 0$, let Q_ε be the conditional probability of

$$P\{X^1[\tau^1_\varepsilon, \tau^1_1] \cap X^2[\tau^2_\varepsilon, \tau^2_1] = \varnothing\}$$

given X^1. Then $p(\varepsilon, j) = E(Q^j_\varepsilon)$. Define a function $b : R^+ \to R$ by

$$b(a) = \lim_{\varepsilon\to 0} \frac{\log P\{Q_\varepsilon \geq \varepsilon^a\}}{\log \varepsilon},$$

i.e., $P\{Q_\varepsilon \geq \varepsilon^a\} \approx \varepsilon^{b(a)}$. It can be shown that the limit exists and that b is a convex function of a. This is closely related to large deviations—if we let

$$L_n = -\log Q_{2^{-n}},$$

then

$$\lim_{n\to\infty} \frac{1}{n} \log P\{L_n \leq (a\log 2)n\} = b(a)\log 2.$$

PROPOSITION 4.1. $\xi(j) = \inf_{a>0}(aj + b(a))$.

The proof of this proposition is not difficult once we know the exponents $\xi(j)$ exist. To see why the RHS is a reasonable quantity, note that if $P\{Q_\varepsilon \geq \varepsilon^a\} \geq \varepsilon^b$, then $p(\varepsilon, j) \geq \varepsilon^{aj+b}$.

Since b is a convex function of a, the infimum is well defined. We expect that b is strictly convex in which case the point at which the infimum is achieved, say a_j, is unique and different for each j. Properties of Brownian motion allow us to derive some properties of $b(a)$, see [2], [3].

If $d = 2$,

- $\lim_{a\to 1/2} b(a) = \infty$ (this is essentially the Beurling projection theorem from harmonic functions, see [1]);
- $b(a) < \infty$ for every $a > 1/2$;
- $\lim_{a\to\infty} b(a) \geq \gamma > 0$, where γ is an exponent defined by

$$P\{X^1[\tau^1_\varepsilon, \tau^1_1] \text{ does not make a closed loop about } 0\} \approx \varepsilon^\gamma.$$

If $d = 3$,

- $\lim_{a\to 0} b(a) = \infty$;
- $b(a) < \infty$ for every $a > 0$;
- $\lim_{a\to\infty} b(a) = 0$.

We now derive inequalities for ξ from the above facts. Let a_1 and a_2 be such that $\xi(1) = a_1 + b(a_1)$, $\xi(a_2) = 2a_2 + b(a_2)$. Then

$$\begin{aligned}\xi(1) &= a_1 + b(a_1)\\ &\leq a_2 + b(a_2)\\ &\leq \xi(2) - a_2.\end{aligned}$$

For $d = 2$, $a_2 > 1/2$, and for $d = 3$, $a_2 > 0$. Therefore

$$\xi_2(1) < \frac{3}{2}; \quad \xi_3(1) < 1.$$

For $d = 2$, we also get $2 = \xi(2) \leq 2a_1 + b(a_1)$. Therefore, $a_1 \geq \frac{1}{2}(2 - b(a_1))$ and

$$\begin{aligned}\xi_2(1) &= a_1 + b(a_1) \\ &\geq 1 + \frac{1}{2}b(a_1) \\ &\geq 1 + \frac{1}{2}\gamma.\end{aligned}$$

Our best estimate of γ is $\gamma \geq \frac{1}{2\pi}$, which gives the best lower bounds of ξ_2 and ζ_2.

REFERENCES

1. L. V. Ahlfors, *Conformal Invariants*, McGraw Hill, 1973.
2. K. Burdzy and G. Lawler, *Non-intersection exponents for Brownian paths, Part* I. *Existence and an invariance principle*, Probab. Theory Related Fields **84** (1990), 393–410.
3. ——, *Nonintersection exponents for Brownian paths, Part* II. *Estimates and applications to a random fractal*, Ann. Probab. **18** (1990), 981–1009.
4. K. Burdzy, G. Lawler, and T. Polaski, *On the critical exponent for random walk intersections*, J. Statist. Phys. **56** (1989), 1–12.
5. B. Duplantier and K.-H. Kwon, *Conformal invariance and intersections of random walks*, Phys. Rev. Lett. **61** (1988), 2514–2517.
6. J. W. Lyklema and C. Evertsz, *Fractals in Physics* (L. Pietronero and E. T. Tossatti, eds.), North Holland, Amsterdam, 1986.
7. H. Kesten, *How long are the arms in* DLA?, J. Phys. A **20** (1987), L29–L33.
8. ——, *Hitting probabilities of random walks on* Z^d, Stoc. Proc. and Appl. **25** (1987), 165–184.
9. G. Lawler, *A self-avoiding random walk*, Duke Math. J. **47** (1980), 655–693.
10. ——, *Intersections of random walks in four dimensions* II, Comm. Math. Phys. **97** (1985), 583–594.
11. ——, *Intersections of simple random walks*, Comm. Math. **41** (1985), 539–554.
12. ——, *Loop-erased self-avoiding random walk in two and three dimensions*, J. Statist. Phys. **50** (1988), 91–108.
13. ——, *Intersections of random walks with random sets*, Israel J. Math. **65** (1989), 113–132.
14. ——, *Estimates for differences and Harnack's inequality for difference equations coming from random walks with symmetric, spatially inhomogeneous increments*, Proc. London Math. Soc. (to appear).
15. F. Spitzer, *Principles of Random Walk*, 2nd ed., Springer-Verlag, New York, 1976.
16. T. A. Witten and L. M. Sander, *Diffusion-limited aggregation, a kinetic critical phenomenon*, Phys. Rev. Lett. **47** (1981), 1400–1403.

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY, DURHAM, NC 27706
E-mail address: jose@math.duke.edu