INDEPENDENT AND DEPENDENT PERCOLATION

J. T. Chayes,† A. Puha and T. Sweet

Department of Mathematics,
University of California,
Los Angeles, CA, 90095

†Partially supported by NSF Grant No. DMS-9403842.

Preface

Percolation is a simple and wonderful model. It is easy to define and yet it exhibits a variety of fascinating phenomena. It is therefore a source of many deep and beautiful mathematical questions.

Independent percolation is the simplest model to undergo a phase transition. Moreover, it has wide application as a model of a random medium. Since its introduction in 1957, independent percolation has been used to describe problems ranging from the distrbution of oil in a porous medium to the distribution of matter in the galaxy. Certain dependent versions of the percolation model, namely the integer random cluster models, are equivalent to some of the basic models of equilibrium statistical mechanics, namely the Ising and Potts magnets.

In the past fifteen years, there has been tremendous progress in the mathematical study of percolation: Stochastic geometric methods have been developed to establish many of the properties of the phase transitions in independent and dependent percolation. These notes review many of these developments.

These notes are based on lectures delivered by one of us (J.T.C.) at the Institute for Advanced Study/Park City Mathematics Institute during the summer of 1996. The first four chapters of the notes draw heavily on lectures given by J.T.C. at the Institute for Mathematics and its Applications Summer School at Ohio State during the summer of 1993, and the first two chapters of these in turn draw on lectures given by J.T.C. and L. Chayes at the Les Houches Summer School in 1984. Much of chapters 5 and 7 of these notes is based (in parts, almost verbatim) on papers of Borgs, Chayes Kesten and Spencer [BCKS] and Borgs and Chayes [BC], respectively. Finally, chapter 6 is based on our understanding of a preliminary manuscript of Reimer; much of the proof in this form was presented by J.T.C. in the Kac Seminars in Utrecht in the summer of 1995.

The contributions of C. Borgs deserve special mention. First, many parts of these notes were based on collaborative efforts between J.T.C. and C. Borgs. In addition, many of the new proofs presented here were derived in collaboration with him. We are indebted to him for these and numerous other contributions.

Table of Contents

Preface 2

1. The Basics of Percolation 4
   1.1 Relevant Quantities and Expected Behavior 4
   1.2 Basic Techniques 14
2. Rescaling and Finite-Size Scaling in Percolation 19
   2.1 Rescaling and Characterization of Phases 19
   2.2 Finite-Size Scaling and the Correlation Length 24
3. Critical Exponent Inequalities 33
   3.1 A Bound on the Correlation Length 33
   3.2 Mean Field Bounds 38

4.Two Fundamental Questions 46
   4.1 Absence of an Intermediate Phase 46
   4.2 Uniqueness of the Infinite Cluster 52

5. Finite-Size Scaling and the Incipient Infinite Cluster 59
   5.1 The Motivation 59
   5.2 Definitions of Relevant Quantities and Preliminaries 62
   5.3 The Scaling Axioms and the Results 65
   5.4 Interpretation of the Results 72
6. The BK(R) Inequality 75
   6.1 Equivalent Forms of the Inequality 75
   6.2 Preliminaries to the Proof of the BK Inequality 77
   6.3 The Proof of the BK Inequality 79
7. The Potts Model and the Random Cluster Model 86
   7.1 The Potts Models 87
   7.2 The Fortuin-Kasteleyn Representation 91
   7.3 Standard Correlation Inequalities 94
   7.4 The BC Decoupling Inequalities 97
   7.5 Comparison Principles 101
   7.6 The DLR Equation and States of the Random Cluster Model 104
   7.7 Length Scales in the Potts Models 110

References 120

CHAPTER I

THE BASICS OF PERCOLATION

We begin these notes with an overview of the percolation model and the tools we will use in our study.

1.1 Relevant Quantities and Expected Behavior

1.1.1 The Model.

The percolation model was introduced in 1957 by Broadbent and Hammersley [BH] to describe the distribution of fluid in a porous medium. There are two basic versions of the model: site or bond percolation, in which the random variables live on either the sites or the bonds of a lattice. In these notes, we will usually discuss bond percolation, although essentially everything we will derive holds for site percolation as well.

Let $\mathbb{L}_d$ denote a regular d-dimensional site lattice, such as the hypercubic lattice $\mathbb{Z}^d$. Let $\mathbb{B}_d$ be the set of bonds between the nearest neighbors of $\mathbb{L}_d$. Then $\Omega = \{0,1\}^{\mathbb{B}_d} = \prod_{b\in\mathbb{B}_d}\{0,1\}$ can be regarded as the set of configurations of lattice bonds. Given a configuration $\omega \in \Omega$ and a bond b, b is said to be *open* or *occupied* if $\omega_b = 1$. Conversely, b is said to be *closed* or *vacant* if $\omega_b = 0$. In the basic version of bond percolation, each bond is open independently of the others with some probability p. Specifically, given $0 \le p \le 1$, μ_b is a measure on $\{0,1\}$ with $\mu_b(1) = p$ and $\mu_b(0) = 1-p$. The probability measure P_p for bond percolation is the Bernoulli product measure:

$$P_p = \prod_{b\in\mathbb{B}_d} \mu_b. \tag{1.1}$$

Let $\mathcal{F}$ be the σ-field of subsets of Ω generated by the finite-dimensional cylinder sets. That is, $\mathcal{F}$ is generated by sets of the form $\{\omega : \omega = \omega' \text{ on } B\}$, where $\omega' \in \Omega$ is a configuration and B is a finite subset of $\mathbb{B}_d$. The probability space of bond percolation is therefore $(\Omega, \mathcal{F}, P_p)$.

There are several obvious ways in which to generalize the above percolation model. First, $\mathbb{B}_d$ need not consist solely of bonds between nearest neighbors; it may include bonds of finite or even infinite range. Second, the probability of a bond being occupied need not be constant over all bonds. That is, p may be a function of b so that $\mu_b(1) = p_b$. This occurs in random systems and also in uniform systems when $\mathbb{B}_d$ includes bonds of infinite range. Finally, the site lattice $\mathbb{L}_d$, and hence the resulting bond lattice $\mathbb{B}_d$, need not be regular. For example, $\mathbb{L}_d$ could be a Poisson distribution of points, in which case the resulting bond lattice is a so-called Vornoi tesselation. We could vary the model even further by doing away with the lattice entirely: the continuum percolation model may be defined on a Poisson distribution of disks or (hyper)spheres. All of these are still independent percolation models (independent because the p_b's are independently distributed), and most of the results we will discuss hold for these models as well. However, in our discussion, we will usually focus on the nearest-neighbor hypercubic bond lattice with uniform occupation probability p.

If, on the other hand, we modify the model so that (1.1) is no longer a product measure—i.e. if the p_b's are no longer independently distributed—then the behavior of the model changes in a fundamental and usually uncontrollable way. However, in special cases, we are still able to make some rigorous statements. One special case, that of the *random cluster measure*, is the subject of Chapter 7.

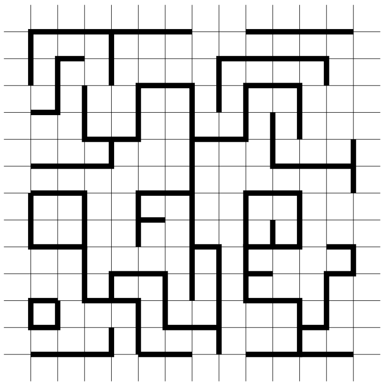


FIGURE 1.1. *A realization of independent bond percolation on the square lattice.*

1.1.2 The Percolation Phase Transition.

A configuration $\omega \in \Omega$ will be identified with the subset of $\mathbb{B}_d$ which is occupied in ω:

$$S(\omega) = \{b \in \mathbb{B}_d : \omega_b = 1\}.$$

$S(\omega)$ consists of disjoint connected components of occupied bonds. We concern ourselves with the typical connectivity properties of $S(\omega)$. For each $x \in \mathbb{L}_d$, let $\mathcal{C}(x) = \mathcal{C}(x;\omega)$ be the set of sites connected to x through occupied bonds. The first question to ask is whether a connected cluster is typically infinite.

Let $|\mathcal{C}(x)|$ denote the number of sites in $\mathcal{C}(x)$. Let $x \leftrightarrow y$, read as *x is connected to y*, be the event that $\mathcal{C}(x) = \mathcal{C}(y)$. Let $x \leftrightarrow \infty$, read as *x is in an infinite cluster*, be the event that $|\mathcal{C}(x)| = \infty$. The quantity

$$P_\infty(p) = P_p(0 \leftrightarrow \infty)$$

is called the *percolation probability* or *infinite cluster density*. If $P_\infty(p) > 0$ (resp. $P_\infty(p) = 0$), then we say that *percolation does (resp. does not) occur*. It is not difficult to see that if percolation exists for p, then percolation also exists for any $p' > p$. So there is a critical threshold which separates the regions where percolation does and does not occur. We say there is a *phase transition*. It is signalled by the nonanalyticity in $P_\infty(p)$, the so-called *order parameter* of the phase transition. We denote that *critical probability* or *percolation threshold* by p_c:

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

We often call the region with $p < p_c$ the *subcritical regime* or *low-density phase*, and the region with $p > p_c$ the *supercritical regime* or *high-density phase*.

REMARKS. (i) If the origin has zero probability of being in an infinite cluster (that is, if $P_\infty(p) = 0$), then there are no infinite clusters anywhere on the lattice with probability 1. Indeed, if $P_\infty(p) = 0$, then

$$\begin{aligned} P_p(\exists x \in \mathbb{L}_d,\ |\mathcal{C}(x)| = \infty) &\le \sum_{x \in \mathbb{L}_d} P_p(|\mathcal{C}(x)| = \infty) \\ &= \sum_{x \in \mathbb{L}_d} P_\infty(p) \\ &= 0, \end{aligned}$$

where the second step follows from translation invariance.

(ii) If the origin has positive probability of being in an infinite cluster, then there is an infinite cluster somewhere on the lattice with probability 1. This is due to Kolmogorov's 0-1 law and the fact that $\{\exists x \in \mathbb{L}_d,\ |\mathcal{C}(x)| = \infty\}$ is a *tail event*. The *tail σ-field*, $\mathcal{F}_\infty$, is defined as one would expect:

$$\mathcal{F}_\infty = \bigcap_B \sigma\Big(\prod_{b \notin B} \mu_b\Big),$$

where the intersection runs over all finite subsets of $\mathbb{B}_d$ and where $\sigma(\nu)$ is the σ-field associated with ν.

(iii) If $P_\infty(p) > 0$, then there is exactly one infinite cluster—provided that the model obeys certain weak conditions to be discussed in more detail in Chapter 4. Unlike the previous two remarks, this is not at all obvious. Indeed, the proof of uniqueness of the infinite cluster is one of the major achievements of percolation theory. In 1981, Newman and Schulman [NS1] proved that with probability one, the number of infinite clusters is either zero, one, or infinity. Then, in 1987, Aizenman, Kesten, and Newman [AKN] showed that the infinite cluster must be unique. In 1989, Burton and Keane [BuK] gave a far easier and more general proof of this uniqueness. We will review the Newman and Schulman and Burton and Keane proofs in Chapter 4.

(iv) The percolation threshold is decreasing in dimension. If there is percolation on the lattice $\mathbb{B}_d \subset \mathbb{B}_{d+1}$, then obviously there is percolation on $\mathbb{B}_{d+1}$. We sometimes write $p_c(d)$ to indicate the explicit dimensional dependence.

(v) For finite-range models in one dimension, the phase transition occurs at the trivial value: $p_c(1) = 1$. This is easy to see, since any density of vacant bonds will disconnect a one-dimensional lattice.

Having made Remark (v), one might wonder if the phase transition occurs at a nontrivial point in any dimension. Indeed it does, as we show below using self-avoiding walk bounds and a Peierls' argument.

THEOREM 1.1. *If* $d \geq 2$, $0 < p_c(d) < 1$.

PROOF OF $0 < p_c(d)$. Here, we will find a $p' > 0$ such that $P_\infty(p) = 0$ for all $p < p'$. To this end, we consider lattice self-avoiding walks (SAW's). Let $\sigma(n)$ be the number of SAW's on $\mathbb{L}_d$ starting at 0 of length n. Let $N(n)$ be the random number of such paths that are occupied. Obviously, the probability of such a path being occupied is p^n. Hence,

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

But, if the origin is part of an infinite cluster, then there must be an occupied SAW of any length emanating from it. Hence, for any n,

$$\begin{aligned} P_\infty(p) &= P_p(|\mathcal{C}(0)| = \infty) \\ &\leq P_p(N(n) \geq 1) \\ &\leq E_p(N(n)) \\ &= p^n \sigma(n). \end{aligned} \tag{1.2}$$

It suffices to give a crude bound on $\sigma(n)$. Let $c(d)$ denote the coordination number, i.e. the number of bonds incident to a site. (For example, $c(d) = 2d$ on $\mathbb{Z}^d$.) Now consider constructing a SAW of length n. First we must choose a bond from the origin; there are $c(d)$ such choices. Then there are only $c(d)-1$ choices for the next bond, since we cannot re-use the previous one. Continuing in this manner, we obtain the bound

$$\sigma(n) \leq c(d)\,(c(d) - 1)^{n-1},$$

which in turn gives us a bound on the percolation probability:

$$P_\infty(p) \leq p\, c(d) \lim_{n\to\infty} [p\,(c(d) - 1)]^{n-1}.$$

Thus if $p < (c(d) - 1)^{-1}$, then $P_\infty(p) = 0$. $\square$

REMARK. Although the above bound suffices to prove the theorem, it is worth noting that it is easy to derive a better lower bound on p_c than $(c(d) - 1)^{-1}$. This follows from the subadditive estimate

$$\sigma(m + n) \leq \sigma(m)\sigma(n).$$

The right hand side above is the number of paths created by "pasting" a SAW of size n to the end of the SAW of size m, which produces, among other things, all SAW's of size $m + n$. By taking logarithms and using a standard subadditivity argument, we have that the limit of $\log(\sigma(n))/n$ exists. (See the proof of Proposition 2.6 for an explicit subadditivity argument.) Thus we may define the *connectivity constant* of the lattice:

$$\lambda = \lambda(d) = \lim_{n\to\infty} \sqrt[n]{\sigma(n)}. \tag{1.3}$$

Using this and (1.2), it follows that $P_\infty(p) = 0$ if $p < \lambda(d)^{-1}$. That is, $\lambda(d)^{-1} \leq p_c$. Note, however, that even this bound is still far short of being sharp.

In order to prove $p_c(d) < 1$, we must first discuss the notion of duality. We begin with $d = 2$. Every planar graph G has a *planar dual* G^* defined as follows:

(1) the vertices of G^* are the centers of the faces of G, and
(2) the bonds of G^* are lines between those vertices of G whose corresponding faces in G have a common edge in their boundaries.

Notice that each bond $b^* \in G^*$ intersects exactly one bond $b \in G$, which we will refer to as the *dual bond* of b^*. We can construct the dual percolation model by declaring that b^* is occupied whenever b is vacant and vice versa. Notice the following geometric fact: Every (occupied) finite connected cluster $C \subset G$ is surrounded by a (occupied) dual circuit $\gamma \subset G^*$. For a concrete example, consider $\mathbb{Z}^2$, which is self-dual: $(\mathbb{Z}^2)^* = (\mathbb{Z} + \frac{1}{2})^2$.

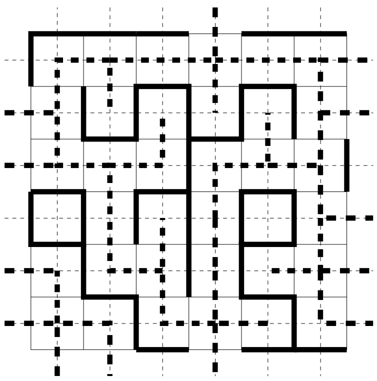


FIGURE 1.2. *Independent bond percolation on $\mathbb{Z}^2$ with density p is dual to bond percolation on $(\mathbb{Z} + \frac{1}{2})^2$ with density $1 - p$. Note that $\mathcal{C}(x)$ is finite if and only if x is contained in the interior of a dual circuit.*

The notion of duality is also useful in higher dimensions. There, bonds are dual to $(d-1)$-cells, and each finite bond cluster on G is surrounded by a dual $(d-1)$-dimensional closed hypersurface.

Using the notion of duality, it is now easy to prove the second half of Theorem 1.1.

PROOF OF $p_c(d) < 1$. Since $p_c(d+1) \leq p_c(d)$, it suffices to show that $p_c(2) < 1$. So, for $d = 2$, we must find a $p' < 1$ such that $P_\infty(p) > 0$ for $p > p'$. We will use a technique known as a *Peierls' estimate*. Let $\rho(n)$ be the number of closed circuits on the dual lattice $\mathbb{B}_2^*$ of length n that enclose 0. Given a dual circuit γ, let A_γ be the event that γ is occupied. Recall that a dual circuit in $\mathbb{B}_2^*$ is occupied if every bond in $\mathbb{B}_2$ that intersects the circuit is vacant. Thus $P(A_\gamma) = (1-p)^n$, where n is the length of γ.

If the origin is not in an infinite cluster, then there must be an occupied dual circuit γ surrounding the origin and conversely. Hence, denoting by $\mathrm{Int}(\gamma)$ the set of points in the original lattice surrounded by the dual circuit γ, we have

$$\begin{aligned} 1 - P_\infty(p) &= P_p(0 \not\leftrightarrow \infty) \\ &= P_p\Big(\bigcup_{\mathrm{Int}(\gamma)\ni 0} A_\gamma \Big) \\ &\leq \sum_{\mathrm{Int}(\gamma)\ni 0} P_p(A_\gamma) \\ &= \sum_n \rho(n)\,(1-p)^n. \end{aligned}$$

To bound $\rho(n)$, consider creating a dual closed circuit of length n surrounding 0. Begin the circuit on a point of the form $(k^*, 0^*)$ with $0 \leq k \leq n$. (This can always be done since the circuit must cross such a point if it is to surround 0 and have length n). Obviously, there are n choices for our starting point. Then, the next $n-1$ steps that we take must form a SAW. Of course, this does not guarantee that we will end up with a circuit, but it does give us an upper bound on $\rho(n)$: $\rho(n) \leq n\,\sigma(n-1)$. We have

$$1 - P_\infty(p) \leq \sum_n n\,\sigma(n-1)\,(1-p)^n. \tag{1.4}$$

Recalling (1.3), the sum in (1.4) must converge for $p > 1 - \lambda(2)^{-1}$, and hence must be strictly less than 1 for p still larger. Hence, $P_\infty(p) > 0$ for p sufficiently close to 1. $\square$

REMARK. Although it seems that the above proof requires that the sum in (1.4) be strictly less than 1, convergence of the sum is actually sufficient to establish $P_\infty(p) > 0$. Indeed, if the sum converges, then it follows that, for some n_0,

$$\sum_{n>n_0} n\,\sigma(n-1)\,(1-p)^n < \delta < 1.$$

Hence, with probability greater than $1-\delta$, there are no occupied dual circuits surrounding 0 of length greater than n_0. But if all bonds inside all such circuits were occupied (which costs no more than $c_1 p^{c_2 n_0^2}$), then with overwhelming conditional probability, the origin would be connected to infinity. If the absence of certain occupied dual circuits and the presence of certain occupied bonds were always positively correlated, then we could conclude that

$$P_\infty(p) \geq (1-\delta)\,\tilde{c}_1 p^{c_2 n_0^2} > 0.$$

That this positive correlation is true follows from the FKG inequality, which will be introduced later in this chapter. Thus $p_c(2) < 1 - \lambda(2)^{-1}$.

1.1.3 Quantities of Interest.

In this subsection, we briefly introduce some of the fundamental quantities in percolation theory. We have already seen $P_\infty(p)$, the percolation probability. It is also referred to as the *order parameter*, since it becomes nonzero at the point at which long-range order appears. $P_\infty(p)$ is analogous to the magnetization in spin models, as will be discussed in some detail in Chapter 7.

A number of other quantities of interest are defined in terms of moments of the cluster size distribution. The expected cluster size is

$$\chi(p) = E_p(|\mathcal{C}(0)|). \tag{1.5}$$

It is analogous to the susceptibility in spin models. Notice that

$$\chi(p) = \sum_{n=1}^{\infty} n\, P_p(|\mathcal{C}| = n) + \infty \cdot P_\infty(p)$$

so that $\chi(p) = \infty$ if $P_\infty(p) > 0$. Using a Peierls' argument, it is easy to show that $\chi(p) < \infty$ for sufficiently small p. Thus, just as $P_\infty(p)$ is used to define the critical value p_c, $\chi(p)$ gives us another critical value:

$$\pi_c = \inf\{p : \chi(p) = \infty\}. \tag{1.6}$$

One of the fundamental results of percolation theory is that $\pi_c = p_c$ on most lattices of interest. In 1980, Kesten [K1] proved the result for the square bond lattice in $d = 2$. Then, in 1986-7, Menshikov, Molchanov, and Sidorenko [MMS] and Aizenman and Barsky [AB] independently proved the result for translation invariant lattices in general dimension. That translation invariance is a necessary condition had been shown earlier by Chayes and Chayes [CC2] who constructed a counterexample on a wedge which opened logarithmically slowly. (See Section 4.1 for a sketch of the Aizenman and Barsky proof and Section 7.5 for a more detailed discussion of the Chayes and Chayes wedges.)

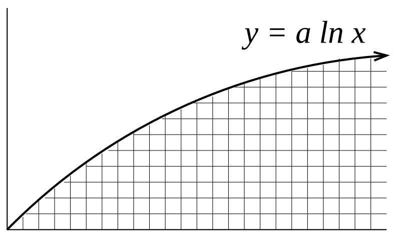


FIGURE 1.3. *$\pi_c < p_c$ for bond percolation on logarithmically opening subsets of the square lattice provided that the constant a is sufficiently small.*

We will also have occasion to consider the expected size of finite clusters. Let

$$P_n(p) = P_p(|\mathcal{C}(0)| = n) \tag{1.7}$$

and

$$P_{\geq n}(p) = P_p(|\mathcal{C}(0)| \geq n). \tag{1.8}$$

It is known that, in the subcritical regime (i.e. when $p < p_c$),

$$P_n(p) \sim e^{-\alpha(p)n}$$

where $f(n) \sim e^{-cg(n)}$ means $c = -\lim_{n\to\infty} g(n)^{-1} \log f(n)$. In the supercritical regime,

$$P_n(p) \sim e^{-\beta(p)n^{(d-1)/d}}$$

which follows from the results of [CCN2], [KZ] and [ADS]. The form of the constant $\beta(p)$ for two-dimensional percolation was determined in [ACC], where it was shown that this constant is related to the shape of large clusters (the so-called Wulff construction for percolation). The asymptotic behavior of $P_n(p)$ implies that the mean size of *finite* clusters is finite for $p \neq p_c$:

$$\chi^{\text{fin}}(p) = E_p(|\mathcal{C}(0)|,\ |\mathcal{C}(0)| < \infty) < \infty. \tag{1.9}$$

A final quantity related to the cluster size distribution is the number of clusters per vertex:

$$\kappa(p) = E_p(|\mathcal{C}(0)|^{-1}). \tag{1.10}$$

Although $0 < \kappa < 1$ for all p, derivatives of $\kappa(p)$ diverge at p_c; one of these derivatives is analogous to the specific heat in spin models.

Next we consider the connectivity function, which is the analogue of the two-point correlation function in spin models. For $x, y \in \mathbb{L}_d$, we define $\tau(x, y; p)$ by

$$\tau(x, y; p) = P_p(x \leftrightarrow y).$$

As we will see later (in Proposition 2.6), for x and y along a coordinate axis, $\tau(x, y; p)$ decays exponentially in $|x - y|$:

$$\tau(x, y; p) \sim e^{-|x-y|/\xi(p)} \tag{1.11}$$

with $\xi(p) < \infty$ whenever $p < p_c$. In the supercritical regime, $\tau(x, y; p)$ is not as interesting since $\xi(p) = \infty$:

$$\begin{aligned}
\tau(x, y; p) &\geq P_p(|\mathcal{C}(x)| = \infty,\ |\mathcal{C}(y)| = \infty) \\
&\geq P_p(|\mathcal{C}(x)| = \infty) P_p(|\mathcal{C}(y)| = \infty) \\
&= P_\infty^2(p),
\end{aligned}$$

where the first step uses uniqueness of the infinite cluster and the second uses the FKG inequality (which we will discuss later in this chapter). In this regime, we instead consider two versions of the truncated connectivity function and note their asymptotic behaviors:

$$\begin{aligned}
\tau^{\text{trun}}(x, y; p) &= \tau(x, y; p) - P_\infty^2(p) \\
&\sim e^{-|x-y|/\xi^{\text{trun}}(p)}
\end{aligned}$$

$$\tau^{\text{fin}}(x,y;p) = P_p(x \leftrightarrow y \not\leftrightarrow \infty) \qquad (1.12)$$
$$\sim e^{-|x-y|/\xi^{\text{fin}}(p)}.$$

While these functions are not the same, their asymptotic behaviors are the same; specifically $\xi_{\text{trun}}(p) = \xi_{\text{fin}}(p) < \infty$ for $p \neq p_c$ [CCGKS]. The decay rates $\xi_(p)$, $\xi_{\text{trun}}(p)$ and $\xi_{\text{fin}}(p) < \infty$ are called *correlation lengths*. They are fundamental quantities which will be considered in some detail in later chapters.

1.1.4 Scaling Theory.

Although the functional dependence of the quantities introduced above depends on details of the model (e.g. the lattice structure), it is widely believed that *in the neighborhood of the critical point*, the functions are independent of these details and depend only on the spatial dimension—and, of course, the fact that this is percolation (rather than some other model, such as the Ising magnet). This robustness is thought to be a general feature of statistical mechanical models with second-order phase transitions. It is known as *universality*. The *univerality class* of a model is presumably determined only by general symmetry properties and the spatial dimension. Models in the same universality class are supposed to have the same critical behavior as characterized by the so-called *critical exponents*. These exponents describe the purported power law behavior of the relevant quantities either approaching or at the critical point.

The approach exponents α, β, γ and ν are defined by

$$\kappa'''(p) \approx |p-p_c|^{-1-\alpha} \qquad \text{as } p \searrow p_c \qquad (1.13)$$
$$P_\infty(p) \approx |p-p_c|^{\beta} \qquad \text{as } p \searrow p_c \qquad (1.14)$$
$$\chi(p) \approx |p-p_c|^{-\gamma} \qquad \text{as } p \nearrow p_c \qquad (1.15)$$
$$\xi(p) \approx |p-p_c|^{-\nu} \qquad \text{as } p \nearrow p_c\,. \qquad (1.16)$$

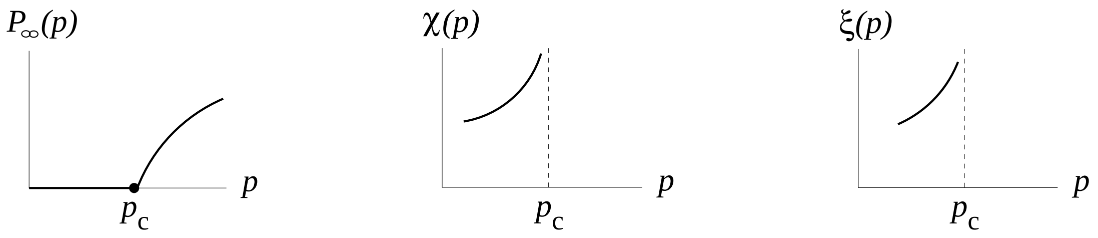


FIGURE 1.4. *The expected behaviors of P_∞, χ, and ξ near p_c.*

One can also define analogues of the exponents γ and ν for approach from above p_c:

$$\chi^{\text{fin}}(p) \approx |p-p_c|^{-\gamma'} \qquad \text{as } p \searrow p_c \qquad (1.17)$$
$$\xi^{\text{fin}}(p) \approx |p-p_c|^{-\nu'} \qquad \text{as } p \searrow p_c\,. \qquad (1.18)$$

It widely believed that $\gamma' = \gamma$ and $\nu' = \nu$; both have been proven in two dimensions, the first in [K4] and the second in [CCGKS].

The exponents δ and η are defined by behaviors at p_c:

$$P_{\geq n}(p_c) \approx n^{-1/\delta} \tag{1.19}$$

and

$$\tau(x, y; p_c) \approx \frac{1}{|x-y|^{d-2+\eta}}\,. \tag{1.20}$$

What is and what is not known rigorously? The first question is whether the system actually has a second-order transition. The defining characteristic is divergence of the correlation length: $\xi(p) \nearrow \infty$ as $p \nearrow p_c$, which is known in all dimensions (see [CC1] and use the fact that $\pi_c = p_c$). It is usually the case that the order parameter, $P_\infty(p)$ is also continuous, i.e. we expect $\lim_{p \searrow p_c} P_\infty(p) = 0$. This is known to be true for the nearest-neighbor model in $d = 2$ [R2], and widely expected—though not known—for finite-range models in higher dimensions. Note, however, that there is a long-range model which has $\lim_{p \searrow p_c} P_\infty(p) > 0$ [ACCN2], although $\lim_{p \nearrow p_c} \xi(p) = \infty$ [IN].

What about the existence of critical exponents such as

$$\lim_{p \searrow p_c} \frac{\log P_\infty(p)}{\log |p - p_c|}?$$

Here we must distinguish what is known in low and high dimension. Most models are expected to have an upper critical dimension, d_c, beyond which the exponents do not change, but instead assume the values they would have in a so-called mean-field model (e.g. on the Cayley tree). The upper critical dimension is thought to be $d_c = 6$ for percolation. Hara and Slade [HS] proved both the existence of critical exponents and the fact that the exponents assume their mean-field values for percolation in sufficiently high dimension. Existence of critical exponents in low dimension is one of the major open problems in percolation theory.

Finally, scaling theory established relations among many of the relevant quantities and therefore among their exponents. Some of the expected scaling relations are

$$\alpha + 2\beta + \gamma = 2,$$

$$\gamma = \nu(2 - \eta)$$

and

$$\beta(\delta - 1) = \gamma.$$

There are also relations involving exponents and the dimension d such as

$$d\nu = 2 - \alpha.$$

Relations involving the dimension are known as hyperscaling relations; they are expected to hold whenever $d \leq d_c$. Kesten [K4] has proved many of the relations among relevant quantities (basically all those not involving κ) in $d = 2$; assuming the existence of the critical exponents, his results imply many of the scaling relations in $d = 2$.

We will return to critical exponents and scaling relations in Chapters 3 and 5. In Chapter 3, we will prove several inequalities on critical exponents (assuming they exist), including a hyperscaling inequality on ν and so-called mean-field inequalities for β and γ, proving that they are always bounded by their mean-field values. In Chapter 5, we will discuss recent results on the critical regime of two-dimensional percolation which rely on the scaling relations of Kesten [K4] and other newly derived scaling relations.

1.2 Basic Techniques

The basic technique of most rigorous work in percolation is the use of correlation inequalities relating, for example, the probability of the intersection of events to the probabilities of the individual events. The simplest correlation inequality is an equality: the factoring of the probabilities of independent events. The two standard nontrivial correlation inequalities of percolation are the Harris-FKG (Fortuin-Kasteleyn-Ginibre) and BK (van den Berg-Kesten) inequalities. These are discussed below, although we postpone a proof of the BK inequality until Chapter 6. In Chapter 7, we discuss also the BC (Borgs-Chayes) decoupling inequalities, which are variants of the FKG inequality that can sometimes substitute for independence and the BK inequality in dependent models.

1.2.1 (FKG) Increasing and Decreasing Events.

There is a natural partial order on the set Ω. Letting $\omega, \omega' \in \Omega$, we say that $\omega \succeq \omega'$ if

$$\omega_b \geq \omega'_b \quad \forall b \in \mathbb{B}_d.$$

Definition: We say that the function (r.v.) $f : \Omega \to \mathbb{R}$ is an *(FKG) increasing function* if it is nondecreasing with respect to this partial order:

$$f(\omega) \geq f(\omega') \quad \forall w \succeq w'.$$

We say that the event $A \in \mathcal{F}$ is an *increasing event* if its indicator function is nondecreasing. Finally, a function g (resp. an event B) is *decreasing* if $-g$ (resp. $\Omega \setminus B$) is increasing.

Informally, increasing events and functions are those that are "helped by the addition of bonds". More formally, we have the following result:

Proposition 1.2. *Let f be an increasing function and let A be an increasing event. Then, for $p_1 \leq p_2$,*

$$E_{p_1}(f) \leq E_{p_2}(f)$$

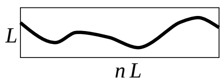
and

$$P_{p_1}(A) \leq P_{p_2}(A).$$

This result is easy to prove and we shall do so shortly. But first we pause to list a few examples of increasing events:

$$\{|\mathcal{C}(0)| = \infty\} = \{0 \leftrightarrow \infty\},$$
$$\{x \in \mathcal{C}(y)\} = \{x \leftrightarrow y\},$$
$$\{|\mathcal{C}(0)| \geq n\},$$

and

$\Bigg\{$ 

$\Bigg\}$.

The last example is the event that there is a long-way crossing by occupied bonds between oppposite faces of an $nL \times L$ rectangle. This "rectangle crossing" event will come up in the next chapter. In addition to the above examples, we wish to add a single "non-example": Notice that

$$\{|\mathcal{C}(0)| = n\}$$

is neither increasing nor decreasing; instead, it is the intersection of increasing and decreasing events.

Now, we return to the proposition above. In the proof, we will use the notation $\omega(b)$ rather than ω_b.

Proof of Proposition 1.2. We define $\bar{\Omega} = [0,1]^{\mathbb{B}_d}$ and define the probability $\bar{P}$ on $\bar{\Omega}$ to be the product of Lebesgue measures; this probability picks uniformly from $[0,1]$ for each bond independently. We now define random variables $w_p \colon \bar{\Omega} \to \Omega$:

$$w_p(b) = w_p(\bar{\omega})(b) = \begin{cases} 1 & \text{if } \bar{\omega}(b) < p \\ 0 & \text{if } \bar{\omega}(b) \geq p. \end{cases}$$

Considering the law of w_p for $0 \leq p \leq 1$, we see that $\bar{P}(w_p \in D) = P_p(D)$. So, we have obtained the bond percolation process on $\mathbb{L}_d$ for all densities on the same probability space. That is, the processes have been *coupled*. This coupling easily yields, for $p_1 \geq p_2$,

$$w_{p_1}(\bar{\omega}) \succeq w_{p_2}(\bar{\omega})$$

for all $\bar{\omega} \in \bar{\Omega}$. So, for any increasing function $f : \Omega \to \mathbb{R}$,

$$f(w_{p_1}(\bar{\omega})) \geq f(w_{p_2}(\bar{\omega})).$$

Hence,

$$\begin{aligned} E_{p_1}(f) &= \bar{E}(f(w_{p_1})) \\ &\geq \bar{E}(f(w_{p_2})) \\ &= E_{p_2}(f). \end{aligned}$$

This is the desired result for functions. For an event A, use the result for the indicator function of A. $\square$

Remark. This coupling in the proof above was used already by Hammersley [Ham2] in the 1960's to simultaneously generate percolation configurations at all $p \in [0,1]$. It was also used extensively by Chayes, Chayes and Newman [CCN1] in their analysis of invasion percolation.

1.2.2 Correlation Inequalities.

The FKG Inequality.

The most widely used correlation inequality is the FKG inequality which says that increasing events are positively correlated. The percolation form of the FKG inequality was proved already in 1960 by Harris [Har]. In 1972, it was formulated and proved for a large class of measures by Fortuin, Kasteleyn and Ginibre [FKG].

THEOREM 1.3. (HARRIS-FKG INEQUALITY) *If A, B are increasing events, then*

$$P_p(A \cap B) \geq P_p(A)P_p(B).$$

REMARKS. (i) Obviously, the inequality is also true if A and B are both decreasing. It reverses direction if one is increasing and the other is decreasing.

(ii) Any measure which satisfies this inequality is called an *FKG measure*.

PROOF OF THEOREM 1.3. We will prove this only for events that depend on the values of a finite number of bonds. The extension to events that depend on an infinite number of bonds (such as $\{|\mathcal{C}(0)| = \infty\}$) is straightfoward. (See for example [Grim2].) In this proof, we will write E and P for E_p and P_p.

It suffices to prove that, if f and g are increasing functions depending on a finite number of bonds, then

$$E(fg) \geq E(f)E(g)\,. \tag{1.21}$$

(1.21) can then be applied to the indicator functions of A and B. The proof of (1.21) is via induction on the number, n, of bonds on which f and g depend.

First, suppose $n = 1$; that is, suppose f and g depend only on the value of ω at a single bond b. So, we think of f and g as functions on $\{0, 1\}$. Let $w, v \in \{0, 1\}$. Then

$$(f(w) - f(v))(g(w) - g(v)) \geq 0$$

since the increasing nature of f and g implies that both factors above are of the same sign. We have

$$\begin{aligned} 0 &\leq \sum_{w,v\in\{0,1\}} (f(w) - f(v))(g(w) - g(v))P(\omega(b) = w)P(\omega(b) = v) \\ &= 2[E(fg) - E(f)E(g)]. \end{aligned}$$

So (1.21) is true when $n = 1$.

Now, suppose (1.21) is true for $n = k$ and let f and g be increasing functions of the values on the $k + 1$ bonds $b(1), \ldots, b(k + 1)$. Let E now denote expectation with respect to the values on those $k + 1$ bonds. Further, let

$$F = E(f|\omega(b(1)), \ldots, \omega(b(k)))$$

and

$$G = E(g|\omega(b(1)), \ldots, \omega(b(k)))\,.$$

Notice that F and G depend only on the first k bond values.

We may evaluate the expectation E on the left hand side of (1.21) in two steps: first taking the expectation with the bond values $b(1), \cdots b(k)$ fixed, and then taking the expectation with respect to $\mu_{b(1)}, \ldots, \mu_{b(k)}$. Since f and g are increasing in the value of the bond $b(k+1)$,

$$\begin{aligned} E(fg) &= E(E(fg|\omega(b(1)), \ldots, \omega(b(k))) \\ &\geq E(FG). \end{aligned}$$

Now since F and G depend on only k bond values, the induction hypothesis tells us that

$$\begin{aligned} E(FG) &\geq E(F)E(G) \\ &= E(f)E(g). \end{aligned}$$

Hence $E(fg) \geq E(f)E(g)$. $\square$

The BK Inequality.

We now have an inequality that bounds the probability of the *intersection* of the two events *below* by the product of their probabilities. Often, we need a complementary inequality which says that the probability of something involving two events is bounded *above* by the product of their probabilities. Obviously, such an inequality cannot be true for the intersection of two events. It turns out that it is true for the *disjoint occurence* of the two events, which we define below.

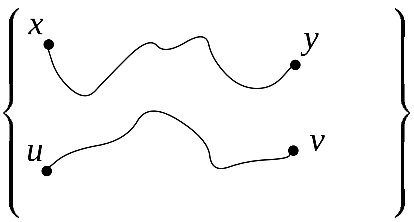
DEFINITION: The event A is said to *occur on the set* S in the configuration ω if if A occurs using only bonds in S, independent of the values of the bonds in S^c. We denote the collection of all such ω by $A|_S$:

$$A|_S = \{\omega : \forall \tilde{\omega},\, \tilde{\omega} = \omega \text{ on } S \Rightarrow \tilde{\omega} \in A\}\,.$$

Two events $A_1, A_2 \in \Omega$ are said to *occur disjointly*, denoted by $A_1 \circ A_2$, if there are two disjoint sets on which they occur:

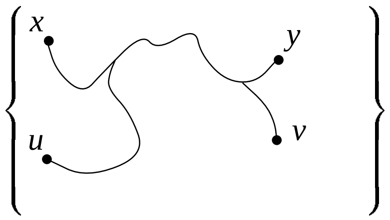
$$A_1 \circ A_2 = \{\omega : \exists S_1, S_2 \subset \mathbb{B}_d,\ S_1 \cap S_2 = \emptyset,\ \omega \in A_1|_{S_1} \cap A_2|_{S_2}\}\,.$$

EXAMPLE: Consider the events $A = \{x \leftrightarrow y\}$ and $B = \{u \leftrightarrow v\}$. It may be that both of these events occur, but on disjoint sets of bonds:



$= A \circ B.$

Alternatively, these two events may occur and may share some common set of bonds:



$$\subset (A \cap B) \setminus (A \circ B).$$

THEOREM 1.4. (BK INEQUALITY) *For all $A, B \in \mathcal{F}$ depending on a finite number of bonds,*

$$P_p(A \circ B) \leq P_p(A)P_p(B).$$

REMARKS. (i) If A is increasing and B is decreasing, then $A \circ B = A \cap B$. So in this case the BK inequality reduces to the FKG inequality.

(ii) The condition that A and B depend on only a finite number of bonds can be relaxed in essentially all examples of interest. See e.g. [Grim2].

The BK inequality was first proved in 1985 by van den Berg and Kesten [BK] for the case in which A and B are both increasing (or both decreasing) events. Then van den Berg and Fiebig [BF] extended it to the case in which A and B are both intersections of increasing and decreasing events. There was subsequently a great deal of work attempting to expand the validity of the inequality to a larger class of events. Finally, a general proof by Reimer [Re] in 1995 confirmed the belief that the inequality holds for all events. This proof will be the subject of Chapter 6.

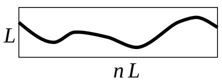
CHAPTER II

RESCALING AND FINITE-SIZE SCALING IN PERCOLATION

2.1 Rescaling and Characterization of Phases ($d = 2$)

In the last chapter, we saw that we could establish the existence of nontrivial high- and low-density phases via Peierls' and SAW arguments. But this only worked when the bond densities were sufficiently large or small. The philosophy in rescaling or real-space renormalization is that, when viewed on a large enough length scale, any noncritical system acts as if it is in a regime of extremely high or low density, often called a Peierls' regime. In particular, except at the critical point, the large-scale analogue of a bond or dual bond is very probable or very improbable.

What is the large scale analogue of a bond? Consider a two-dimensional rectangle of height L and length nL. Say that a *left-right crossing* occurs if there is an open path connecting the left side of the rectangle to the right side of the rectangle. Let $R_{n,L}(p)$ be the probability that there is a left-right crossing of the rectangle at bond density p. That is,

$R_{n,L}(p) = P_p \Big\{$ 

$\Big\}.$

For simplicity, we write $R_{1,L} = R_L$ since in the case $n = 1$, the rectangle is actually a square of side length L.

In this subsection, we will show that, whenever $p > p_c$, $R_{n,L}(p)$ tends to 1 exponentially in the scale L—i.e., it acts as if it is in a Peierls' regime. We will then use this to give a characterization of the high-density phase of two-dimensional percolation. Then, in the next subsection, we will show how the dual of $R_{n,L}(p)$ can be used to define an alternative to the conventional correlation length, $\xi(p)$, in the low-density phase of percolation in any dimension.

2.1.1 Two Lemmas.

We begin with two lemmas which allow us to rescale the bond variables. The first, often known as the RSW lemma, allows us to get from a square of side L to a $2L \times L$ rectangle. It is a strictly two-dimensional lemma. The second, often called the rescaling or ACCFR lemma, shows us how to change the basic scale L. The ACCFR lemma is easily generalizable to higher dimensions, provided that we consider not bonds, but dual $(d-1)$-cells—see remark (iii) following the statement of Lemma 2.3.

Lemma 2.1. (Russo-Seymour-Welsh [R1], [SeW]) *Let $d = 2$. There is a continuous, increasing function $F : [0,1] \to [0,1]$ such that $F(0) = 0$, $F(1) = 1$ and*

$$R_{2,L}(p) \geq F(R_L(p)).$$

REMARKS. (i) In simple terms, the RSW lemma says that if the probability of crossing a square is sufficiently large, then the probability of crossing a $2L \times L$ rectangle—the hard way—is also reasonably large.

(ii) The proof of the RSW lemma is complicated to write down and very particular to two dimensions. We will not go over it here. The reader is referred to [R1] or Grimmett [Grim2] for a complete argument. However, there is an easy and useful proposition which is used three times in the proof of the RSW lemma and which is worth seeing. We note, however, that this proposition is not used in the remainder of these notes.

PROPOSITION 2.2. (THE SQUARE ROOT PROPOSITION) *Suppose A_1 and A_2 are increasing events and $P(A_1) = P(A_2)$. Then*

$$P(A_1) \geq 1 - \sqrt{1 - P(A_1 \cup A_2)}.$$

PROOF. $(A_1 \cup A_2)^c = A_1^c \cap A_2^c$. Hence,

$$\begin{aligned} 1 - P(A_1 \cup A_2) &= P(A_1^c \cap A_2^c) \\ &\geq P(A_1^c)P(A_2^c) \\ &= (1 - P(A_1))^2, \end{aligned}$$

where we have used the FKG inequality in the second step and the assumption $P(A_1) = P(A_2)$ in the final step. $\square$

LEMMA 2.3. (AIZENMAN, CHAYES, CHAYES, FRÖHLICH, RUSSO [ACCFR]) *Suppose $d = 2$. Let $c = \frac{1}{16}$ and $\lambda \in (0, 1)$. If*

$$R_{2,L}(p) \geq 1 - c\lambda,$$

then

$$R_{2,2L}(p) \geq 1 - c\lambda^2.$$

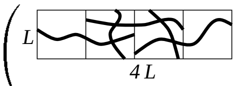
COROLLARY 2.4. *If $R_{2,L}(p) \geq 1 - c\lambda$, then $R_{2,2^kL}(p) \geq 1 - c\lambda^{2^k}$.*

REMARKS. (i) The corollary tells us that if the crossing probability is large enough, then, looked at on larger length scales, it tends to the trivial high-density fixed point ($p = 1$) exponentially fast.

(ii) How do we know whether the crossing probability is large enough to satisfy the hypothesis of the ACCFR lemma? The hypothesis obviously is not satisfied for $p < p_c$. For $p > p_c$, we will show that $\liminf_{L\to\infty} R_L(p) = 1$. Then the RSW lemma allows us to get from an $L \times L$ box to a $2L \times L$ box to show that $\liminf_{L\to\infty} R_{2,L}(p) = 1$. Hence, we will be able to find an L to satisfy the ACCFR lemma.

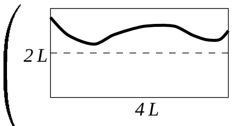
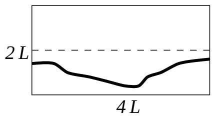
(iii) The analogous lemma holds if we replace $R_{2,L}$ by $R_{n,L}$ with any $n \geq 2$ provided c is chosen accordingly; that is, if $R_{n,L}(p) \geq 1 - c\lambda$, then $R_{n,2L}(p) \geq 1 - c\lambda^2$. More interestingly, an analogous lemma also holds in higher dimensions (with a different value of c) for long-way crossings of $(d-1)$-dimensional *hypersurfaces*. Note, however, that the lemma does not hold for *bond* crossings in $d > 2$; this is because the intersections required in the proof can be ensured only for objects of codimension 1.

PROOF OF LEMMA 2.3. We have

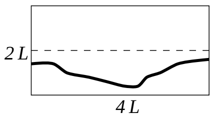
$$R_{4,L}(p) \geq P_p\Big($$

$$\Big)$$

$$\begin{aligned} &\geq R_{2,L}(p)^3 R_L(p)^2 \\ &\geq R_{2,L}(p)^4 \\ &\geq (1-c\lambda)^4 \\ &\geq 1-4c\lambda, \end{aligned} \tag{2.1}$$

where the second inequality follows from the FKG inequality, the third comes from the fact that any crossing of a $2L \times L$ rectangle can be decomposed into two separate crossing of disjoint $L \times L$ rectangles, and the fourth is the assumption of the lemma. But

$$R_{2,2L}(p) \geq P_p\Big($$

or

$$\Big), \tag{2.2}$$

so that, by independence of the two crossings above,

$$1 - R_{2,2L}(p) \leq P_p\Big(\Big\{$$

$$\Big\}^c\Big)^2$$

$$\begin{aligned} &= (1-R_{4,L}(p))^2 \\ &\leq (4c\lambda)^2 \\ &= c\lambda^2, \end{aligned}$$

where have used the estimate (2.1) in the second to last step, and the fact that $c = \frac{1}{16}$ in the final step. □

Notice that the key estimate in the proof above is the *choice* in (2.2) which leads to the *squaring* of the probability.

2.1.2 Characterization of the High-Density Phase.

The following theorem shows that square crossings give a characterization of the entire high-density phase of two-dimensional percolation.

THEOREM 2.5. *In $d = 2$, $p > p_c$ if and only if* $\liminf_{L\to\infty} R_L(p) = 1$.

PROOF. Assume $\liminf_{L\to\infty} R_L(p) = 1$. Then, by RSW, there exists an L_0 such that

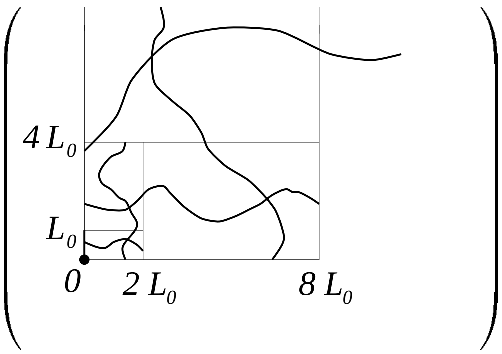
$$R_{2,L_0}(p) \geq 1 - \frac{c}{e} \tag{2.3}$$

with $c = \frac{1}{16}$. By the corollary to the ACCFR lemma, this in turn implies

$$R_{2,2^k L_0}(p) \geq 1 - \frac{c}{e^{2^k}}. \tag{2.4}$$

Note that our choice of $\lambda = 1/e$ in the ACCFR Lemma was arbitrary; any $\lambda < 1$ would suffice. Using the FKG inequality and the above estimate, it is now easy to construct an infinite cluster from the origin. Indeed,

$$P_\infty(p) \geq P_p$$



$$\begin{aligned}
&\geq p^{L_0} \prod_{k=1}^{\infty} R_{2,2^k L_0} \\
&\geq p^{L_0} \prod_{k=1}^{\infty} \left(1 - \frac{c}{e^{2^k}}\right) \\
&> 0\,.
\end{aligned}$$

This proves the "if" part of the theorem.

Now assume that $p > p_c$. Let us first show that $R_L(p) > 0$. We will want to consider the probability that the origin connects to the boundary of a box of side L centered at the origin. We decompose this event by recording exactly which portion of the boundary connects to the origin. For this purpose, we define the face i of ∂B_L, denoted by $F_i(\partial B_L)$, as

$$F_i(\partial B_L) = \{x \in \partial B_L : x_{|i|} = \mathrm{sign}(i)\ L\}$$

for $i \in \{1, -1, \ldots, d, -d\}$. In $d = 2$,

$$\begin{aligned}
&P_p(0 \leftrightarrow \partial B_L) \\
&\qquad = P_p(0 \leftrightarrow F_1(\partial B_L) \text{ or } 0 \leftrightarrow F_{-1}(\partial B_L) \text{ or } 0 \leftrightarrow F_2(\partial B_L) \text{ or } 0 \leftrightarrow F_{-2}(\partial B_L)) \\
&\qquad \leq 4P_p(0 \leftrightarrow F_1(\partial B_L))\,,
\end{aligned} \tag{2.5}$$

where we have used subadditivity of the measure and rotation invariance in the last step. Since $P_p(0 \leftrightarrow \partial B_L) \geq P_\infty(p)$, (2.5) implies that

$$P_p(0 \leftrightarrow F_1(\partial B_L)) \geq \frac{1}{4} P_\infty.$$

Pasting together a connection from the origin to face 1 and a connection from the origin to face -1 gives a horizontal crossing of the $L \times L$ square. Therefore

$$\begin{aligned} R_L(p) &\geq P_p(0 \leftrightarrow F_1(\partial B_L) \ \cap \ 0 \leftrightarrow F_{-1}(\partial B_L)) \\ &\geq P_p(0 \leftrightarrow F_1(\partial B_L))^2 \\ &\geq \frac{1}{16} P_\infty^2(p) \\ &> 0\,. \end{aligned} \tag{2.6}$$

Next we want to show that as L gets larger, $R_L(p)$ actually tends to 1. We do this with a coarse-grained version of the above argument. Let $\epsilon > 0$. Since $p > p_c$, there is an infinite cluster with probability 1, and hence there exists an N large enough such that

$$P_p(\partial B_N \leftrightarrow \infty) \geq 1 - \epsilon\,. \tag{2.7}$$

Take $L \gg N$ and construct a coarse-grained analogue of $P_p(0 \leftrightarrow \partial B_L)$ by replacing the vertex at the origin with a box of size N. That is, consider $P_p(\partial B_N \leftrightarrow \partial B_L)$. We have

$$\begin{aligned} P_p(\partial B_N \not\leftrightarrow \partial B_L) &= P_p(\partial B_N \not\leftrightarrow F_i(\partial B_L) \text{ for all } i = 1, -1, 2, -2) \\ &\geq P_p(\partial B_N \not\leftrightarrow F_1(\partial B_L))^4 \end{aligned} \tag{2.8}$$

by the FKG inequality and rotational invariance. Since

$$P_p(\partial B_N \not\leftrightarrow \infty) \geq P_p(\partial B_N \not\leftrightarrow \partial B_L)\,,$$

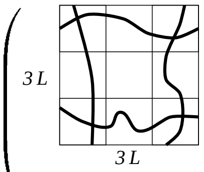
(2.7) implies

$$P_p(\partial B_N \not\leftrightarrow \partial B_L) \leq \epsilon$$

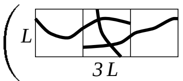
which, together with (2.8), gives the bound

$$P_p(\partial B_N \not\leftrightarrow F_1(\partial B_L)) \leq \epsilon^{1/4}. \tag{2.9}$$

Let $A_{N,L}$ denote the event that there exists an open circuit in the annulus $B_L \setminus B_N$. We have

$$P_p(A_{L,3L}) \geq P_p \left(\vphantom{\int} \right.$$

$$\left. \vphantom{\int} \right)$$

$$\geq (R_{3,L}(p))^4$$

$$\geq \left(P_p \left(\vphantom{\int} \right. \right.$$

$$\left. \left. \vphantom{\int} \right) \right)^4$$

$$\begin{aligned} &\geq \left((R_{2,L}(p))^2 \, R_L(p) \right)^4 \\ &\geq (F(R_L(p)))^8 \, R_L(p)^4 \\ &\geq \left(F(\frac{1}{16} P_\infty^2) \right)^8 \left(\frac{1}{16} P_\infty^2 \right)^4 \\ &\equiv a > 0, \end{aligned} \tag{2.10}$$

where we have used FKG and rotational invariance to get the second and fourth inequalities, RSW for the fifth, and (2.6) for the sixth. The fact that $a > 0$ uniformly in L follows from $p > p_c$. In order to obtain a lower bound on $P_p(A_{N,L})$, we partition $B_L \setminus B_N$ into a sequence of nonoverlapping annuli: $A_1 = B_{3N} \setminus B_N$, $A_2 = B_{9N} \setminus B_{3N}$, and so on. Notice that there are $n = n(L) = \lfloor (\log L - \log N)/\log 3 \rfloor$ independent annuli in $B_L \setminus B_N$. If any one of these annuli contains an open circuit, then so does $B_L \setminus B_N$. Hence,

$$P_p(A_{N,L}) \geq P_p(\bigcup_{i=1}^{n(L)} A_{3^{i-1}N, 3^i N}).$$

Using BK on the complements and (2.10) gives

$$1 - P_p(A_{N,L}) \leq (1-a)^{n(L)}. \tag{2.11}$$

Now let $\delta > 0$. Then, by (2.11), there exists L_0 such that for every $L > L_0$

$$P_p(A_{N,L}) \geq 1 - \delta. \tag{2.12}$$

Finally, we note that the $L \times L$ square crossing can be ensured by having two crossings from the inner $N \times N$ square to opposing faces of the $L \times L$ square, together with a circuit in the annulus $A_{N,L}$. Thus

$$\begin{aligned} R_L(p) &\geq P_P(\partial B_N \leftrightarrow F_1(\partial B_L),\ \partial B_N \leftrightarrow F_{-1}(\partial B_L),\ A_{N,L}) \\ &\geq (1-\epsilon^{1/4})^2(1-\delta) \end{aligned}$$

where the final inequality comes from FKG, (2.9), and (2.12). Choosing ϵ and δ appropriately makes this bound as close to 1 as desired. $\square$

It is worth noting that we can substitute the estimate of the above proof into the RSW and ACCFR lemmas to show that, in fact, $R_L(p)$ tends to 1 exponentially fast. By the ACCFR lemma, we need only ensure that

$$R_{2,L}(p) > \frac{15}{16},$$

which, by the RSW lemma, in turn requires

$$R_L(p) > F^{-1}(\frac{15}{16}),$$

where F is the function defined in the RSW lemma.

2.2 Finite-Size Scaling and the Correlation Length

We have just shown that on large enough length scales, the system appears to be in a Peierls' regime. But what is "large enough"? According to our analysis above, "large enough" means large enough to satisfy the conditions of the rescaling lemma. Thus, for two dimensions, we define

$$L_0^*(p) = \min\{L \geq 1 : R_{2,L}(p) \geq 1 - \frac{c}{e}\}, \tag{2.13}$$

or in general

$$L_0^*(p; n, \epsilon) = \min\{L \geq 1 : R_{n,L}(p) \geq 1 - \epsilon\}\,, \tag{2.14}$$

where $c, \epsilon > 0$ and n are chosen to satisfy the conditions of the ACCFR rescaling lemma. The choice of $1/e$ in equation (2.13) is simply for computational convenience. For $L > L_0^*(p)$, we are effectively in a high-density Peierls' regime, in the sense that our effective bond density tends exponentially to 1. Thus, in some sense, $L_0^*(p)$ is the high-density length scale beyond which the system acts noncritically.

One of the fundamental, but rather vague and certainly unproven, assumptions of scaling theory is that there is only one length scale in the system: the *correlation length*—or, for $p > p_c$, the dual correlation length. Thus we would expect that $L_0^*(p)$ should be equivalent, in some appropriate sense, to the two-dimensional high-density (dual) correlation length. This was, in fact, the original motivation for the definition of a finite-size scaling correlation length by Fröhlich, Chayes and Chayes [CCF]. [CCF] called $L_0^*(p)$ a finite-size scaling length because it is equivalent to the dual correlation length and it is determined by the scaling of the probabilities of events which occur in *finite* volumes. The notion of a finite-size scaling correlation length was later used by Chayes, Chayes, Fisher and Spencer [CCFS1] in their analysis of the correlation length exponent (see Section 3.1), by Kesten [K4] in his analysis of critical scaling relations, and by Borgs, Chayes, Kesten and Spencer [BCKS] in their analysis of the critical regime. The [BCKS] work will be the subject of Chapter 5, where we will discuss these notions in more detail.

For the purposes of this chapter, let us simply note that it is easier to make these notions precise in the subcritical regime of percolation than in the supercritical regime. Thus, we will first show that the fundamental correlation length for $p < p_c$—i.e. the inverse decay rate of $\tau(x, y; p)$—is well-defined. Then we will define an analogue of $L_0^*(p)$ below p_c and shows that, up to logarithms, this analogue is equivalent to the fundamental correlation length. These results will hold for percolation in all dimensions.

2.2.1 The Fundamental Correlation Length for $p < p_c$.

The first question to address is existence of the fundamental correlation length. Recall

$$\tau(x, y; p) = P_p(x \leftrightarrow y).$$

We will define the correlation length in terms of the on-axis connectivity. To this end, let e_i be the unit vector in the i^{th} direction.

PROPOSITION 2.6. (EXISTENCE OF A FUNDAMENTAL CORRELATION LENGTH) *The limit*

$$\lim_{n\to\infty} \frac{\log \tau(0, ne_1; p)}{n} := -\frac{1}{\xi(p)}$$

exists.

PROOF. Fix $m < n$. We have

$$\begin{aligned}\tau(0, ne_1) &\geq P_p(0 \leftrightarrow me_1 \leftrightarrow ne_1)\\ &\geq \tau(0, me_1)\,\tau(me_1, ne_1)\\ &= \tau(0, me_1)\,\tau(0, (n-m)e_1),\end{aligned} \tag{2.15}$$

where the second inequality is an application of FKG and the final equality follows from translation invariance. Whenever a quantity obeys an inequality of the form (2.15), we will say that it is *subadditive.* (Note, however, that technically $\tau(0, ne_1; p)$ is supermultiplicative, and only $-\log(\tau(0, ne_1; p))$ is subadditive.)

The subadditive inequality (2.15) is enough to prove existence of the limit. Choose k_n such that

$$\lim_{n\to\infty} \frac{\log(\tau(0, k_n e_1))}{k_n} = \liminf_{n\to\infty} \frac{\log(\tau(0, ne_1))}{n} := -\frac{1}{\xi}.$$

Fix a nonnegative integer m. Define

$$\ell_n = \max\{\ell : \ell m \leq k_n\} \quad \text{and} \quad r_n = k_n - \ell_n m.$$

Iterating (2.15), we have

$$\begin{aligned}\tau(0, k_n e_1) &\geq \tau(0, \ell_n m e_1)\,\tau(0, r_n e_1)\\ &\geq \tau(0, me_1)^{\ell_n}\,\tau(0, r_n e_1).\end{aligned}$$

Taking logarithms and dividing both sides by k_n gives

$$\frac{\log(\tau(0, k_n e_1))}{k_n} \geq \frac{\ell_n m}{k_n}\frac{\log(\tau(0, me_1))}{m} + \frac{\log(\tau(0, r_n e_1))}{k_n}.$$

Noting that $\ell_n m/k_n$ tends to 1 and the remainder term tends to 0, we have

$$-\frac{1}{\xi} \geq \frac{\log(\tau(0, me_1))}{m}. \tag{2.16}$$

Since m is chosen arbitrarily, we may take lim sup of (2.16) to obtain

$$\lim_{m\to\infty} \frac{\log(\tau(0, me_1))}{m} = -\frac{1}{\xi}. \quad \square$$

From (2.16), we immediately have the *a priori bound*:

COROLLARY 2.7. (THE A PRIORI BOUND)

$$\tau(0, me_1; p) \leq \exp(-\frac{m}{\xi(p)}) \tag{2.17}$$

Next one would like to know that $\xi(p)$ is not trivial (i.e., that $\xi(p) \neq \infty$). In order to prove this, we will use an inequality which is often called the Lieb-Simon inequality after Lieb [L] and Simon [S] who proved the analogue of it for Ising systems in 1980. A percolation version of it was proved by much earlier by Hammersley [Ham1].

THEOREM 2.8. (HAMMERSLEY-LIEB-SIMON INEQUALITY) *Let B be a box centered at the origin and ∂B its boundary. For $x \in \partial B$, let $\tau'(0, x; p) \leq \tau(0, x; p)$ denote the probability of a connection between 0 and x using only occupied bonds within B. Then*

$$\tau(0, z; p) \leq \sum_{x \in \partial B} \tau'(0, x; p)\, \tau(x, z; p). \tag{2.18}$$

PROOF. Let $\mathcal{C}^B(0) \subset \mathcal{C}(0)$ denote the connected component of the origin within B. Let

$$T(0, z) = \{0 \leftrightarrow z\}$$

and

$$T'(0, z) = \{0 \leftrightarrow z \text{ in } B\},$$

so that $\tau(0, z; p) = P_p(T(0, z))$ and $\tau'(0, z; p) = P_p(T'(0, z))$. If $z \notin B$, then, in order for the origin to connect to z, there must be an $x \in \partial B$ such that $x \in \mathcal{C}^B(0)$ and a path from x to z in the complement of $\mathcal{C}^B(0)$.

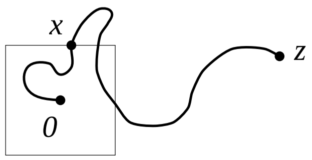


FIGURE 2.1. *Notice the path from $x \in C^B(0)$ to z in the complement of $C^B(0)$.*

Therefore, we can write

$$T(0, z) = \{\omega : \exists x \in \partial B \text{ such that } \omega \in T'(0, x) \circ T(x, z)\}$$

or, equivalently,

$$T(0, z) = \bigcup_{x \in \partial B} T'(0, x) \circ T(x, z).$$

Using the BK inequality, we have

$$P_p(T'(0, x) \circ T(x, z)) \leq P_p(T'(0, x))\, P_p(T(x, z))$$

and therefore,

$$\tau(0, z; p) \leq \sum_{x \in \partial B} \tau'(0, x; p)\, \tau(x, z; p). \quad \square$$

THEOREM 2.9. *If $p < p_c$, then $\xi(p) < \infty$.*

PROOF. Notice that, for every ω,

$$|\mathcal{C}(0;\omega)| = \sum_{x\in\mathbb{Z}^d} \mathbf{1}_{T(0,x)}(\omega),$$

where $\mathbf{1}_A$ is the indicator function of the event A. Taking expected values gives

$$\chi(p) = E_p(|\mathcal{C}(0)|) = \sum_{x\in\mathbb{Z}^d} \tau(0,x;p). \tag{2.19}$$

Since $p_c = \pi_c$, choosing $p < p_c$ implies that $\chi(p) < \infty$, and thus that the series on the right hand side of (2.19) converges. Now

$$\sum_{x\in\partial B} \tau'(0,x) \le \sum_{x\in\partial B} \tau(0,x),$$

which, by (2.19), is bounded by the tail of a convergent series. Hence, for B sufficiently large,

$$\sum_{x\in\partial B} \tau'(0,x) < 1.$$

Define $\alpha > 0$ by

$$e^{-\alpha} = \sum_{x\in\partial B} \tau'(0,x).$$

Then, for $x \in \partial B$,

$$W(x) = e^{\alpha}\tau'(0,x)$$

is a normalized set of weights.

Now pick z far from B and rewrite the HLS inequality (2.18) as

$$\tau(0,z;p) \le e^{-\alpha} \sum_{x\in\partial B} W(x)\,\tau(x,z;p). \tag{2.20}$$

Translation invariance implies that

$$\tau(x,z;p) \le e^{-\alpha} \sum_{y\in\partial B(x)} W(y-x)\,\tau(y,z;p) \tag{2.21}$$

where $B(x)$ is the box B translated by x. Rewriting (2.20) by inserting (2.21) in place of $\tau(x,z;p)$ gives

$$\tau(0,z;p) \le e^{-2\alpha} \sum_{x\in\partial B(0)} W(x) \sum_{y\in\partial B(x)} W(y-x)\,\tau(y,z;p).$$

This substitution can be iterated $n = \lfloor \frac{|z|}{\mathrm{diam}(B)} \rfloor$ times and still guarantee that z is in the complement of all of the translates of B.

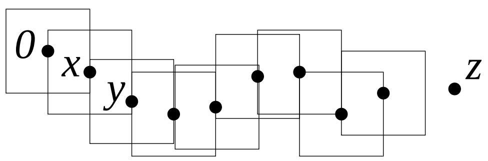


FIGURE 2.2. *Notice that the n^{th} iterate of the process creates a point at most $n\,diam(B)$ units from the origin.*

After iterating this n times, we have

$$\begin{aligned}
&\tau(0,z;p)\\
&\le \exp\left(\lfloor\frac{-\alpha|z|}{\operatorname{diam}(B)}\rfloor\right)\sum_{x_1\in\partial B(0)}W(x_1)\sum_{x_2\in\partial B(0)}W(x_2)\cdots\sum_{x_n\in\partial B(0)}W(x_n)\,\tau(x_n,z)\\
&\le \exp\left(\lfloor\frac{-\alpha|z|}{\operatorname{diam}(B)}\rfloor\right).
\end{aligned}$$

The final inequality follows from the facts that $\tau(x_n,z)\le 1$ and that the weights are normalized. Choose $z_k = ke_1$. Taking logarithms, dividing by k, and letting k tend to infinity, we obtain

$$\frac{1}{\xi(p)} \ge \frac{\alpha}{\operatorname{diam}(B)} > 0. \quad \square$$

REMARK. Notice that we do not actually use the prime of the HLS inequality (2.18) in the proof of Theorem 2.9. We bound τ' from above by τ and use the fact that for $p<p_c$, $\sum_x \tau(0,x)$ is convergent. Hammersley [Ham1] proved a version of (2.18) without the prime for percolation, as did Simon [S] for the Ising magnet. Lieb [L] added the prime to Simon's Ising magnet inequality; this version was used to prove that $\xi(p)$ diverges *continuously* in the Ising model. The crucial point in the proof was that τ' is the probability of an event which occurs in a *finite* box. For percolation, we can prove continuity of $\xi(p)$ by other finite-size considerations. We follow closely the proof of Proposition 2.11 in [CC1].

THEOREM 2.10. $\lim_{p\to p_c}\xi(p)=\infty$.

PROOF. First we sketch a proof that $\xi^{-1}(p)$ is left continuous, i.e.

$$\lim_{\epsilon\searrow 0}\xi^{-1}(p-\epsilon)=\xi^{-1}(p)$$

for all p. To this end, we define yet another connectivity function, $\tau^T(0,x)$, as the probability that 0 is connected to x by a path of occupied bonds that lies entirely in the "tunnel" $\{x: -T\le x_2,\cdots,x_d\le T\}$. Let $(\xi^{-1})^T(p)$ denote the corresponding inverse correlation length. It is straightforward to show that:
(i) $\forall T$, $(\xi^{-1})^T(p)$ is continuous and decreasing in p, and
(ii) $\forall p$, $(\xi^{-1})^T(p)\searrow\xi^{-1}(p)$ as $T\nearrow\infty$.
(See [CC1], remark 3 following the proof of Theorem 5.1, for an analogous proof for SAWs.) So $\xi^{-1}(p)$ is a decreasing limit of a sequence of continuous, decreasing functions and is therefore left continuous.

Next we show that $\xi^{-1}(p)$ approaches zero *continuously* (which, of course, does not follow from left continuity). This is done by a variant of the argument used by Russo [R2] to show continuity of $P_\infty(p)$ at p_c in $d = 2$. Suppose that, for some p_0, $\xi^{-1}(p_0) > 0$. We must show that, for ϵ sufficiently small, $\xi^{-1}(p_0 + \epsilon) > 0$. To this end, note that $\xi^{-1}(p_0) > 0$ implies there is some L_0 so that the probability that a hypersheet of dual $(d-1)$-cells separates the left from the right faces of a $L_0 \times 2L_0 \times 2L_0 \times \cdots \times 2L_0$ box exceeds $1 - c(d)/e$, where $c(d)$ is the constant in the d-dimensional ACCFR rescaling lemma. But, since the box is finite, this probability is an analytic (indeed, a polynomial) function of p, so there is an ϵ small enough so that the probability of the hypersurface crossing event still exceeds $1 - c(d)/e$ at $p = p_0 + \epsilon$. But now we can paste these rectangles together with ACCFR rescaling (see e. g. the proof of Theorem 2.11) to show that $\xi^{-1}(p_0 + \epsilon) > 0$, as desired.

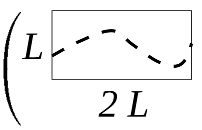
Thus we have that $\xi^{-1}(p)$ approaches zero continuously. Now, given Theorem 2.9, we need only show that $\xi^{-1}(p_c) = 0$. Suppose the opposite: $\xi^{-1}(p_c) > 0$. Then, by the construction discussed above, we can rescale hypersurfaces in boxes at density $p_c + \epsilon$. Intersecting these hypersurfaces, we can create closed surfaces in hyperannuli with probability tending to one exponentially in the size of the hyperannuli. These closed surfaces can then be used to obtain a finite upper bound on $\chi(p_c + \epsilon)$, a contradiction. □

2.2.2 The Finite-Size Scaling Correlation Length.

Just as we can define a two-dimensional finite-size scaling length, $L_0^*(p)$, for $p > p_c$ as the smallest length above which renormalized bonds will rescale, we can define a finite-size scaling correlation length, $L_0(p)$, for $p < p_c$ as the smallest length above which renormalized dual bonds (or hypersurfaces) will rescale. Since, as noted in remark (iii) following Lemma 2.3, in $d > 2$, dual hypersurfaces rescale (i.e. satisfy an ACCFR rescaling lemma), whereas bonds do not, the finite-size scaling correlation length so defined will make sense in all dimensions.

In d dimensions, let $R_{n,L}(p)$ denote the probability of an occupied bond crossing in the e_1 direction of an $nL \times L \times L \times \cdots \times L$ box. Similarly, let $R^*_{n,L}(p)$ be the probability that an $L \times nL \times nL \times \cdots \times nL$ box is spanned (the hard way) by an unbroken hypersurface of dual $(d-1)$-cells, so that, in particular, $R^*_{n,L} = 1 - R_{1/n,nL}$. Schematically,

$$R^*_{2,L}(p) = P_p \left(\qquad \right),$$



where the dotted line represents the dual hypersurface. Following [CCF], we define the *finite-size scaling correlation length* $L_0(p)$ for $p < p_c$ as

$$L_0(p) = \min\{L \geq 1 : R^*_{2,L}(p) \geq 1 - \frac{c}{e}\}, \tag{2.22}$$

or, in general,

$$\begin{aligned} L_0(p; n, \epsilon) &= \min\{L \geq 1 : R^*_{n,L}(p) \geq 1 - \epsilon\} \\ &= \min\{L \geq 1 : R_{1/n,nL}(p) \leq \epsilon\}. \end{aligned} \tag{2.23}$$

Here, c is the dimension-dependent constant in the rescaling lemma, e.g. $c = \frac{1}{16}$ in $d = 2$ (see Lemma 2.3 and remark (iii), following it). In general ϵ and n are chosen to satisfy the hypothesis of the d-dimensional ACCFR rescaling lemma. The next theorem show that, up to logarithms, $\xi(p)$ and $L_0(p)$ coincide.

THEOREM 2.11. (CHAYES, CHAYES AND FRÖHLICH [CCF]) *Let $p < p_c$. There exist nonzero, finite, dimension-dependent constants c_1 and c_2 such that*

$$\frac{c_1 \log L_0(p) + c_2}{L_0(p)} \geq \frac{1}{\xi(p)} \geq \frac{1}{L_0(p)}.$$

REMARKS. (i) The theorem implies that if $\xi(p)$ and $L_0(p)$ both diverge with power laws as $p \nearrow p_c$, then their critical exponents must coincide. See equation (1.16).

(ii) Recently, K. Alexander has shown that, in $d = 2$, one may take $c_1 = 0$ in the above theorem.

(iii) The theorem also holds for $L_0(p; n, \epsilon)$ rather than $L_0(p)$ if n and ϵ are chosen to satisfy the hypothesis of the d-dimensional ACCFR Lemma. Now, however, c_1 and c_2 depend on n and ϵ.

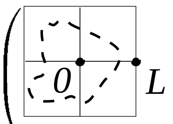
(iv) Analogous results for $p > p_c$ and $d = 2$ are discussed in Chapter 5.

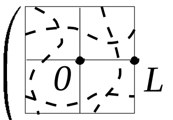
PROOF OF THEOREM 2.11. We begin by proving the lower bound on $1/\xi$. To this end, let $L = 2^k L_0$. By d-dimensional ACCFR rescaling and the definition of L_0, we have

$$R^*_{2,L} \geq 1 - c\exp(-2^k) = 1 - c\exp(-L/L_0). \tag{2.24}$$

Now, by obvious bounds and the FKG inequality,

$$1 - \tau(0, Le_1) = P(0 \not\leftrightarrow Le_1)$$

$\geq P_p\Bigg($  $\Bigg)$

$\geq P_p\Bigg($  $\Bigg)$

$$\geq \left(R^*_{2,L}\right)^{2d}$$

which, together with (2.24), gives the estimate

$$\begin{aligned} \tau(0, Le_1) &\leq 1 - \left(R^*_{2,L}\right)^{2d} \\ &\leq 1 - (1 - c\exp(-L/L_0))^{2d} \\ &\leq \exp(-L/L_0)\,. \end{aligned}$$

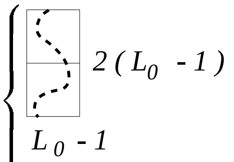
Taking logarithms, dividing by $-L$, and letting L tend to infinity, we get

$$\frac{1}{\xi} \geq \frac{1}{L_0},$$

which proves the lower bound.

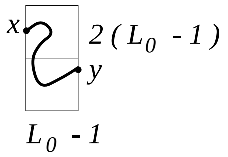
Next we prove the upper bound on $1/\xi$. We have

$$1 - R^*_{2,L_0-1} = P_p\Bigg(\Bigg\{$$



$$\Bigg\}^c\Bigg)$$

$$= P_p\Bigg(\bigcup_{x,y}\Bigg\{$$



$$\Bigg\}\Bigg)$$

$$\le \sum_{x\in F_{-1}} \sum_{y\in F_1} \tau(x,y)$$

$$\le (2L_0)^{2(d-1)} \exp\left(-\frac{(L_0-1)}{\xi}\right), \tag{2.25}$$

where the second step is a result of duality, the third step follows from subadditivity of the measure and relaxing the constraint that the connection occur within the box, and the final step is an application of the *a priori* bound (2.17). Now recall that, by definition, L_0 is the smallest L such that $R^*_{2,L}(p) \ge 1 - \frac{c}{e}$. Hence,

$$R^*_{2,L_0-1} \le 1 - \frac{c}{e},$$

which, by (2.25), implies that

$$\begin{aligned} \frac{c}{e} &\le 1 - R^*_{2,L_0-1} \\ &\le (2L_0)^{2(d-1)} \exp\left(-\frac{(L_0-1)}{\xi}\right) \end{aligned}$$

and therefore

$$\log\left(\frac{c}{2^{2(d-1)}e}\right) \le 2(d-1)\log L_0 - \frac{L_0-1}{\xi}.$$

Rewriting this gives an upper bound of

$$\frac{1}{\xi} \le \frac{2(d-1)\log L_0 + \left|\log\left(\frac{c}{2^{2(d-1)}e}\right)\right|}{L_0-1}.$$

Obviously, the values of the constants c_1 and c_2 depend on the choice of n and ϵ, but otherwise the proof is essentially identical for any n and ϵ which satisfy the hypothesis of the rescaling lemma. $\square$

CHAPTER III

CRITICAL EXPONENT INEQUALITIES

In the last chapter, we saw that for $p \neq p_{\mathrm{c}}$, the system can be characterized by the existence of a length scale beyond which connectivities decay exponentially. At many multiples of this scale, the system is essentially in a Peierls' regime—either low or high density. What characterizes a system with a second-order transition is that at the transition point, the model has no length scale—the system is said to be critical or *scale-invariant*. In Chapter 5, we will discuss new theorems which quantitatively describe the nature of critical regime in percolation. Here we will focus on the exponents which describe the approach to the critical regime. In particular, we will prove bounds on the exponents ν, β and γ, provided these exponents exist.

3.1 A Bound on the Correlation Length

In Theorem 2.10, we saw that percolation has a second order transition in the sense that $\xi(p)$ diverges continuously as p increases to p_{c}. Now we will show more—we will get a lower bound on the rate of divergence, which therefore gives us a bound on the exponent ν, provided it exists. First, we need a probabilistic lemma.

LEMMA 3.1. (CHAYES, CHAYES, FISHER, SPENCER [CCFS1]) *Let A be an event which depends only on bonds in a finite set Λ. Take $p \in (0,1)$. Then*

$$\left|\frac{dP_p(A)}{dp}\right| \leq \alpha(p)\sqrt{|\Lambda|}$$

where $\alpha(p) = 1/\sqrt{p(1-p)}$ and $|\Lambda|$ denotes the number of bonds in Λ.

PROOF. We have

$$\begin{aligned} P_p(A) &= \sum_{\omega \in \Omega_\Lambda} P_p(\omega)\mathbf{1}_A(\omega) \\ &= \sum_{\omega \in \Omega_\Lambda} p^{n(\omega)}(1-p)^{|\Lambda|-n(\omega)}\mathbf{1}_A(\omega) \end{aligned}$$

where $\Omega_\Lambda = \{0,1\}^\Lambda$ and $n(\omega)$ is the number of occupied bonds in the configuration

$\omega \in \Omega_\Lambda$. Differentiating this expression gives

$$\begin{aligned}
&\frac{\mathrm{d}P_p(A)}{\mathrm{d}p} \\
&= \sum_{\omega\in\Omega_\Lambda} \left(n(\omega)p^{n(\omega)-1}(1-p)^{|\Lambda|-n(\omega)} - (|\Lambda| - n(\omega))\, p^{n(\omega)}(1-p)^{|\Lambda|-n(\omega)-1} \right) \mathbf{1}_A(\omega) \\
&= \sum_{\omega\in\Omega_\Lambda} \left(\frac{n(\omega)}{p} - \frac{|\Lambda| - n(\omega)}{1-p} \right) P_p(\omega)\mathbf{1}_A(\omega).
\end{aligned}$$

Note that

$$\begin{aligned}
\frac{n(\omega)}{p} - \frac{|\Lambda| - n(\omega)}{1-p} &= \frac{n(\omega) - p\,|\Lambda|}{p(1-p)} \\
&= \frac{n(\omega) - E_p(n)}{p(1-p)},
\end{aligned}$$

and hence,

$$\frac{\mathrm{d}P_p(A)}{\mathrm{d}p} = \frac{1}{p(1-p)} \sum_{\omega\in\Omega_\Lambda} P_p(\omega)\,(n(\omega) - E_p(n))\,\mathbf{1}_A(\omega).$$

We therefore have the bound

$$\left|\frac{\mathrm{d}P_p(A)}{\mathrm{d}p}\right| \leq \frac{1}{p(1-p)} \sum_{\omega\in\Omega_\Lambda} P_p(\omega)\,|n(\omega) - E_p(n)|\,.$$

Using the Cauchy-Schwarz inequality and the fact that n is binomial $(|\Lambda|\,, p)$, we finally obtain

$$\begin{aligned}
\left|\frac{\mathrm{d}P_p(A)}{\mathrm{d}p}\right| &\leq \frac{1}{p(1-p)} \left(\sum_{\omega\in\Omega_\Lambda} P_p(\omega)\,(n(\omega) - E_p(n))^2 \right)^{\frac{1}{2}} \\
&= \frac{1}{p(1-p)} \sqrt{\mathrm{Var}_p(n)} \\
&= \frac{1}{p(1-p)} \sqrt{p(1-p)\,|\Lambda|} \\
&= \frac{1}{\sqrt{p(1-p)}} \sqrt{|\Lambda|},
\end{aligned}$$

the desired bound. $\square$

Finite-Size Scaling Events and Correlation Lengths

Lemma 3.1 tells us that the change in $P_p(A)$ due to a change in the bond density is no worse than a square-root-of-volume fluctuation. Next, we will see how this square-root-of-volume fluctuation leads to a lower bound on ν. To this end, consider the σ-algebra of events which depend only on bonds in a finite box Λ, with $|\Lambda| = L^d$. We denote this algebra by $\mathcal{F}_L$ and say that $A_L \in \mathcal{F}_L$ is a *finite-size scaling event*. Given a sequence $A_L \in \mathcal{F}_L$ and a constant $\tilde{c} > 0$, we define the *finite-size scaling correlation length*

$$\xi_{FSS}(p) = \xi_{FSS}(p; A_L, \tilde{c}) = \max\{L \geq 1 : P_p(A_L) > \tilde{c}\}. \tag{3.1}$$

Notice that if $\lim_{L\to\infty} P_p(A_L) = 0$, then $\xi_{FSS}(p) < \infty$.

We will see in the proof of Corollary 3.3 below that our percolation finite-size scaling correlation length (2.22) is bounded below by a length of the form (3.1). However, the definition above applies under much more general conditions. For example, it is possible to construct a finite-size scaling length of the form (3.1) for dilute and random Ising ferromagnets, and to show that this length is equivalent to the fundamental correlation length defined as the so-called quenched average of the two-point correlation function [CCFS2]. The Thouless length, which is used extensively in the physics literature as a correlation length in disordered electronic systems, is also of the form (3.1), although in this case, it is not yet possible to show that this length is equivalent to a more fundamental correlation length. In any case, the generality of the definition (3.1) means that the following theorem is applicable to many disordered systems in addition to bond percolation.

THEOREM 3.2. ([CCFS1]) *Let $A_L \in \mathcal{F}_L$, $\tilde{c} > 0$ and $\xi_{FSS}(p) = \xi_{FSS}(p; A_L, \tilde{c})$. Suppose there exists L_1 such that for all $L > L_1$, $P_{p_c}(A_L) \geq 2\tilde{c}$. Then*

$$\xi_{FSS}(p) \geq c(p)\,(p_c - p)^{-\frac{2}{d}}\,, \tag{3.2}$$

where $c(p)$ is bounded away from 0 in a neighborhood of p_c.

COROLLARY 3.3. *If $\xi(p)$ is the correlation length defined in Proposition 2.6, then*

$$\liminf_{p\nearrow p_c} \frac{\log(\xi(p))}{|\log(p_c - p)|} \geq \frac{2}{d}. \tag{3.3}$$

REMARK. If the exponent ν defined in (1.16) exists, then the corollary implies that it satisfies $\nu \geq \frac{2}{d}$.

PROOF OF COROLLARY 3.3. Recall that $R^*_{2,L}(p)$ is the probability that an $L \times 2L \times \cdots \times 2L$ box is spanned by an unbroken hypersurface of dual $(d-1)$-cells. Take c to be the constant such that $R^*_{2,L} \geq 1 - c/e$ implies the rescaling of such surfaces (see the remark (iii) following Lemma 2.3). By duality, $R^*_{2,L} = 1 - R_{\frac{1}{2},2L}$ where $R_{\frac{1}{2},2L}$ is the probability of a short-way bond crossing of an $L \times 2L \times \cdots \times 2L$ box. Recalling the definition (2.22), we have

$$\begin{aligned} L_0(p) &= \min\{L \geq 1 : R^*_{2,L}(p) \geq 1 - \frac{c}{e}\} \\ &= \min\{L \geq 1 : R_{\frac{1}{2},2L}(p) \leq \frac{c}{e}\}\,. \end{aligned} \tag{3.4}$$

In particular,

$$L_0(p) - 1 \le \max\{L \ge 1 : R_{\frac{1}{2},2L}(p) > \frac{c}{e}\}\,. \tag{3.5}$$

Thus, if we could show that $R_{\frac{1}{2},2L}$ was monotone decreasing in L, we could conclude that

$$L_0(p) = \max\{L \ge 1 : R_{\frac{1}{2},2L}(p) > \frac{c}{e}\}(+1)$$

where $a = b(+1)$ means a equals b or $b+1$. This would then imply that $L_0(p)$ is of the form of the finite-size scaling correlation length defined in (3.1) with A_L the event of a short-way occupied bond crossing of an $L \times 2L \times \cdots \times 2L$ box and $\tilde{c} = c/e$.

Unfortunately, however, we do not know whether $R_{\frac{1}{2},2L}$ is monotone in L, so we require an additional argument. Roughly speaking, the argument says that although $R_{\frac{1}{2},2L} = P(A_L)$ is not necessarily monotone, it is essentially monotone in the sense that

$$P(A_L) \le P(A_{L_0}) \qquad \text{for all } L \ge 2L_0\,. \tag{3.6}$$

First, let us show that this is enough to bound $L_0(p)$ from below by a length of the form (3.1), and then let us show that (3.6) is actually satisfied. By equation (3.4), we have $P(A_{L_0}) \le \tilde{c}$. Putting this together with (3.6), we conclude that

$$P(A_L) \le P(A_{L_0}) \le \tilde{c} \qquad \text{for all } L \ge 2L_0.$$

But by the definition (3.1) of $\xi_{FSS}(p)$, we have

$$P(A_{\xi_{FSS}}) > \tilde{c}\,.$$

Putting together the last two equations with (3.5), we conclude

$$L_0(p) - 1 \le \xi_{FSS}(p) < 2L_0(p)\,,$$

as desired.

Now, to complete the argument, let us verify equation (3.6). To this end, we claim that a rescaling argument along the lines of the proof of the ACCFR lemma gives the general result

$$P(A_{2L+k}) \le C(d) P(A_L)^2 \tag{3.7}$$

for all nonnegative $k \le 2L$. Indeed, let $Q_{N,M}$ denote the probability of a bond crossing in the e_1 direction of an $N \times M \times M \cdots \times M$ box, so that $P(A_L) = R_{\frac{1}{2},2L} = Q_{L,2L}$. For any nonnegative $k \le 2L$, the fact that $Q_{N,M}$ is decreasing in N and increasing in M implies

$$Q_{2L+k,2(2L+k)} \le Q_{2L,8L}\,.$$

Observing that a bond crossing of a $2L \times 8L \times 8L \cdots \times 8L$ box implies two independent bond crossings of $L \times 8L \times 8L \cdots \times 8L$ boxes, we have

$$Q_{2L+k,2(2L+k)} \le Q_{L,8L}^2\,. \tag{3.8}$$

Notice that this last step is analogous to equation (2.2) in the proof of the rescaling lemma. Now doing a surface patching argument along the lines of equation (2.1), we obtain

$$1 - Q_{L,8L} \geq (1 - Q_{L,2L})^{\sqrt{C(d)}} \geq 1 - \sqrt{C(d)} Q_{L,2L}\,, \tag{3.9}$$

where $\sqrt{C(d)}$ is an integer patching constant. Equations (3.8) and (3.9) imply the general rescaling relation (3.7). To show that (3.7) implies (3.6), we iterate (3.7) ℓ times, obtaining

$$P(A_{2^\ell L+k}) \leq [C(d)P(A_L)]^{2^\ell - 1} P(A_L) \tag{3.10}$$

for all $\ell \geq 1$ and $0 \leq k \leq 2^\ell L$. Now, by choosing $\tilde{c}$ in the definition of L_0 small enough, we may assume that $C(d)P(A_{L_0}) \leq 1$. Substituting this into (3.10) gives

$$P(A_{2^\ell L_0+k}) \leq P(A_{L_0}) \tag{3.11}$$

for all $\ell \geq 1$ and $0 \leq k \leq 2^\ell L_0$. But this is equivalent to the desired bound (3.6).

Finally, we must show that the probability of the finite-size scaling event, $R_{\frac{1}{2},2L} = P(A_L)$, satisfies the hypothesis of Theorem 3.2. First, note that there is a constant $a > 0$ such that

$$R_{\frac{1}{2},2L}(p_c) > a$$

uniformly in L. Otherwise, $L_0(p)$ would not diverge as $p \nearrow p_c$, and hence, by Theorem 2.11, neither would $\xi(p)$, a contradiction. Moreover, by choosing a smaller $\tilde{c}$ if necessary, we may take $a = 2\tilde{c}$, so that the hypothesis is of the theorem is satisfied. (In order to take $a = 2\tilde{c}$, we must actually define $L_0(p) = L_0(p;\tilde{c})$ for different values of the constant $\tilde{c}$ and show that these are equivalent in the sense of Theorem 2.11, but this is no problem.) Thus, by Theorem 3.2, $L_0(p)$ obeys the inequality (3.2). But, by Theorem 2.11, up to logarithms, $\xi(p)$ is bounded above and below by L_0, so that $\xi(p)$ satisfies (3.3). □

Proof of Theorem 3.2. By Lemma 3.1,

$$\left|\frac{\mathrm{d}P_p(A_L)}{\mathrm{d}p}\right| \leq \alpha(p)\sqrt{|\Lambda|} = \alpha(p)L^{\frac{d}{2}}.$$

Thus

$$|P_p(A_L) - P_{p_c}(A_L)| = \left|\int_p^{p_c} \frac{\mathrm{d}P_p(A_L)}{\mathrm{d}p}\mathrm{d}p\right| \leq \int_p^{p_c} \left|\frac{\mathrm{d}P_p(A_L)}{\mathrm{d}p}\right| \mathrm{d}p \leq (p_c - p)\tilde{\alpha}(p)L^{\frac{d}{2}}$$

where

$$\tilde{\alpha}(p) = \frac{1}{p_c - p}\int_p^{p_c} \sqrt{\frac{1}{s(1-s)}}ds\,.$$

This in turn implies

$$\begin{aligned} P_p(A_L) &\geq P_{p_c}(A_L) - (p_c - p)\tilde{\alpha}(p)L^{\frac{d}{2}} \\ &\geq 2\tilde{c} - (p_c - p)\tilde{\alpha}(p)L^{\frac{d}{2}}\,, \end{aligned}$$

where the second step holds for $L > L_1$ by the assumption of the theorem. Taking

$$p = p_c - \frac{\tilde{c}}{\tilde{\alpha}(p)} L^{-\frac{d}{2}} \tag{3.12}$$

gives

$$P_p(A_L) \geq \tilde{c}.$$

Since $\xi_{FSS}(p)$ is defined as the largest L such that $P_p(A_L) \geq \tilde{c}$, it follows that

$$\xi_{FSS}(p) \geq L = \left(\frac{\tilde{\alpha}(p)}{\tilde{c}}\right)^{-\frac{2}{d}} (p_c - p)^{-\frac{2}{d}},$$

where the equality is a consequence of (3.12). The desired result now follows from the fact that $\tilde{\alpha}(p)$ is bounded away from 0 near p_c. □

How sharp is the result $\nu \geq 2/d$? For two-dimensional percolation, it is widely believed that $\nu = 4/3$, which compares rather favorably with the bound $\nu \geq 1$. The result is actually much sharper for some other disordered systems. As mentioned previously, it was shown in [CCFS2] that $\nu \geq 2/d$ also applies to dilute and random Ising ferromagnets. For dilute magnets in three dimensions, the bound $\nu \geq 2/3$ is actually quite close to the best known values (see [CCFS1] and the references therein). For disordered electronic systems, the bound is, in some sense, even sharper. In 1986, when $\nu \geq 2/d$ was first derived, the accepted value of ν for disordered electronic systems in three dimensions was roughly half of the lower bound. The bound showed that there were problems in the physical theory used to interpret experimental data on disordered electronic systems (again, see [CCFS1] and the references therein).

3.2 Mean-Field Bounds

We can also get inequalities on other exponents (i.e., γ, β, δ), but in contrast to our inequality $\nu \geq 2/d$, the inequalities on these other exponents are independent of the dimension. For several statistical mechanical models, it is known that when the dimension becomes high enough, the models become simpler in the sense that they have the same critical exponents as one finds on the Cayley tree or on the complete graph. It turns out that many critical exponents always have values either above or below their high-dimensional constant values. For example, $\gamma \geq 1$ and $\beta \leq 1$ for all d. These bounds are called *mean-field bounds*.

3.2.1 Pivotal Bonds and Russo's Formula.

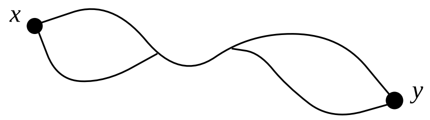
First we introduce the notion of the *pivotal, or articulation, bonds*. Roughly speaking, a bond b is pivotal for an event A in configuration ω if changing the value of ω at b (but at no other bond) changes whether the event A occurs. In other words, pivotal bonds are essential for the event. Formally:

DEFINITION: Let $b \in \mathbb{B}_d$, $\omega \in \Omega$, and define $\omega^{b,+}$ or $\omega^{b,-}$ to be configurations agreeing with ω on $\mathbb{B}_d \setminus \{b\}$ with b occupied or vacant, respectively. We say that b *is pivotal for A in ω* if

$$\mathbf{1}_A(\omega^{b,+}) - \mathbf{1}_A(\omega^{b,+}) \neq 0.$$

The set of bonds which are pivotal for A in ω will be denoted by $\delta A(\omega)$, and its size by $|\delta A(\omega)|$. The event $b \in \delta A$, sometimes denoted by $\delta_b A$, is determined by the σ-algebra on the complement of b.

EXAMPLE: Consider the following configuration in $T(x, y) = \{x \leftrightarrow y\}$:



The bonds in the two circuits are not pivotal for $T(x, y)$ in ω, but the bonds in the path joining the circuits are pivotal.

Not surprisingly, it turns out that the change in probability of an increasing event as a function of bond density is related to the expected number of pivotal bonds. The relation expressing this is known as Russo's formula [R2], although it was actually proved first by Margulis [Ma]. A nice discussion of Russo's formula and related issues may be found in the paper by Friedgut and Kalai [FrK]. The proof given below is due to Chayes and Chayes [CC1].

LEMMA 3.4. (RUSSO'S FORMULA [Ma], [R2]) *Let A be an increasing event which depends only on a finite number of bonds. Then*

$$\frac{dP_p(A)}{dp} = E_p(|\delta A|).$$

REMARK. This can easily be extended to an *in*equality on the right derivative when A depends on an infinite number of bonds (see [CC4]):

$$\frac{\mathrm{d}P_p(A)}{\mathrm{d}p} \equiv \lim_{\varepsilon \searrow 0} \frac{P_{p+\varepsilon}(A) - P_p(A)}{\varepsilon} \geq E_p(|\delta A|). \tag{3.13}$$

PROOF. Let Λ be the finite set on which A depends. Let p_b be the probability that bond b is open. We have

$$\frac{\mathrm{d}P_p(A)}{\mathrm{d}p} = \sum_{b \in \Lambda} \frac{\partial}{\partial p_b} P_p(A)|_{p_b = p}.$$

Fix $b \in \Lambda$ and define $\Omega_{\Lambda \setminus b} = \{0, 1\}^{\Lambda \setminus b}$. We have

$$P_p(A) = \sum_{\omega^b \in \Omega_{\Lambda \setminus b}} P_p(\omega^b) \left(p_b \mathbf{1}_A(\omega^{b,+}) + (1 - p_b) \mathbf{1}_A(\omega^{b,-}) \right),$$

which implies that

$$\frac{\partial P_p(A)}{\partial p_b} = \sum_{\omega^b \in \Omega_{\Lambda \setminus b}} P_p(\omega^b) \left(\mathbf{1}_A(\omega^{b,+}) - \mathbf{1}_A(\omega^{b,-}) \right).$$

Since A is increasing, we have

$$\mathbf{1}_A(\omega^{b,+}) - \mathbf{1}_A(\omega^{b,-}) = \mathbf{1}_{\{b\in\delta A\}}(\omega^b).$$

Thus

$$\frac{\partial P_p(A)}{\partial p_b} = \sum_{\omega^b\in\Omega_{\Lambda\setminus b}} P_p(\omega^b)\mathbf{1}_{\{b\in\delta A\}}(\omega^b) \;=\; P_p(b\in\delta A),$$

and

$$\frac{\mathrm{d}P_p(A)}{\mathrm{d}p} = \sum_{b\in\Lambda} P_p(b\in\delta A) = E_p(|\delta A|). \quad \square$$

3.2.2 A Bound on the Percolation Probability.

Here we will show how the infinite cluster can be decomposed into a *backbone* and *dangling ends*. We will then use this decomposition to prove a differential inequality on the percolation probability, which in turn implies a mean-field bound on the exponent β. In the next chapter, we will show how a generalization of this differential inequality can be used to prove absence of an intermediate phase in percolation. The differential inequality here was proved by Chayes and Chayes ([CC3], [CC4]); its generalization was proved by Aizenman and Barsky [AB].

THEOREM 3.5. ([CC3], [CC4]) *Let $p\in(0,1)$. Then*

$$P_\infty \;\le\; P_\infty^2 \;+\; pP_\infty\frac{dP_\infty}{dp}. \tag{3.14}$$

Taking $p>p_c$, dividing by P_∞, and integrating, we immediately have:

COROLLARY 3.6. *For $p>p_c$,*

$$P_\infty(p) \;-\; \frac{p_c}{p}P_\infty(p_c) \;\ge\; \frac{p-p_c}{p}.$$

REMARK. Assuming $P_\infty(p_c)=0$, which is known to be true in $d=2$ [R2] and is widely believed to be true in all dimensions, and the existence of the exponent β defined in (1.14), the corollary implies that $\beta\le 1$.

In order to prove the theorem, we introduce the notion of the *backbone* of the infinite cluster and the concept of a *spineless* infinite cluster.

DEFINITIONS: A vertex $x\in\mathbb{Z}^d$ is said to belong to the *backbone* of the infinite cluster if it is connected to infinity by two (bond) disjoint paths of occupied bonds. A *spineless* infinite cluster is an infinite cluster with no backbone (e.g. an infinite cluster consisting exclusively of the positive x-axis). We use the following notation for these objects:

$$\begin{aligned}
\mathcal{P}_\infty(0) &= \{0\leftrightarrow\infty\}\\
P_\infty(p) &= P_p(\mathcal{P}_\infty(0))\\
\mathcal{B}_\infty(0) &= \{0\leftrightarrow\infty \text{ along two bond disjoint occupied paths}\}\\
B_\infty(p) &= P_p(\mathcal{B}_\infty(0))\\
\mathcal{S}_\infty(0) &= \{0\leftrightarrow\infty \text{ and the cluster is spineless}\}\\
S_\infty(p) &= P_p(\mathcal{S}_\infty(0)).
\end{aligned}$$

It turns out that we do not have to worry about spineless clusters because configurations containing them are of measure zero:

LEMMA 3.7. ([CC4]) *For* $p > p_c$, $S_\infty(p) = 0$.

We will not bother to prove this here except to say that the proof shows that if $S_\infty(p) > 0$, then P_∞ is discontinuous at p. This contradicts a result of [AKN] which says that $P_\infty(p)$ is continuous for all $p > p_c$. The reader is referred to [CC4] for the complete argument. Given the absence of spineless clusters, we can now prove the differential inequality:

PROOF OF THEOREM 3.5. If the origin is in the infinite cluster, then it is either in the backbone or not. If it is not in the backbone, we say that it is in a *dangling end* of the infinite cluster. Writing the event that the origin is in the infinite cluster as the disjoint union of the backbone and the dangling end events,

$$\mathcal{P}_\infty(0) \;=\; \mathcal{B}_\infty(0) \cup (\mathcal{P}_\infty(0) \setminus \mathcal{B}_\infty(0)),$$

and taking probabilities, we have

$$P_\infty(p) \;=\; B_\infty(p) \;+\; P_p(\mathcal{P}_\infty(0) \setminus \mathcal{B}_\infty(0)). \tag{3.15}$$

Now by the obvious observation

$$\mathcal{B}_\infty(0) = \mathcal{P}_\infty(0) \circ \mathcal{P}_\infty(0)$$

and the BK inequality, we have

$$B_\infty(p) \;\le\; P_\infty^2(p).$$

Hence, we have reduced the problem to obtaining the desired bound on the dangling end term.

Since $S_\infty = 0$ by Lemma 3.7, it follows that, with probability one, the dangling end event $\mathcal{P}_\infty(0) \setminus \mathcal{B}_\infty(0)$ implies the existence of a unique bond $b \in \mathbb{B}_d$ such that
(i) b has an endpoint in the backbone, and
(ii) $b \in \delta\mathcal{P}_\infty(0)$ (i.e., b is pivotal for $\mathcal{P}_\infty(0)$).
This fact is rather obvious geometrically (see Figure 3.1) . The reader is referred to [CC4] for an explicit proof. Let $\mathcal{D}_b(0)$ be the event that $b \in \mathbb{B}_d$ is this unique bond. Then

$$\mathcal{P}_\infty(0) \setminus \mathcal{B}_\infty(0) = \bigcup_{b \in \mathbb{B}_d} \mathcal{D}_b(0).$$

Observe that $\mathcal{D}_b(0)$ requires that b be pivotal for the event $\mathcal{P}_\infty(0)$, and that, in addition, b must be occupied and an endpoint of b must be connected to infinity. Denoting the latter two events by $\Theta(b)$ and $\mathcal{P}_\infty(\partial b)$, respectively, we have

$$\mathcal{D}_b(0) \;=\; \delta_b\mathcal{P}_\infty(0) \,\circ\, \Theta(b) \,\circ\, \mathcal{P}_\infty(\partial b). \tag{3.16}$$

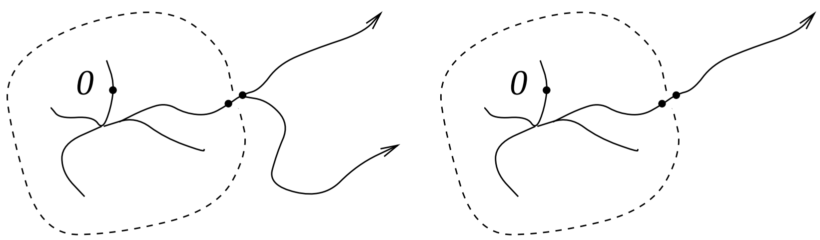


FIGURE 3.1. *Two realizations of $\delta_b P_\infty(0)$. The first realization is in $\mathcal{D}_b(0)$, while the second is not.*

Using equation (3.16), the BK inequality and Russo's formula in the form (3.13), we have

$$\begin{aligned} P_p(\mathcal{P}_\infty(0) \setminus \mathcal{B}_\infty(0)) &\le \sum_{b \in \mathbb{B}_d} P_p(\mathcal{D}_b(0)) \\ &\le \sum_{b \in \mathbb{B}_d} P_p(\delta_b \mathcal{P}_\infty(0)) p P_\infty(p) \\ &= p P_\infty(p) \sum_{b \in \mathbb{B}_d} P_p(\delta_b \mathcal{P}_\infty(0)) \\ &\le p P_\infty(p) \frac{\mathrm{d}P_\infty}{\mathrm{d}p}. \end{aligned} \tag{3.17}$$

Equations (3.15) and (3.17) imply the desired result:

$$P_\infty \le P_\infty^2 + p P_\infty \frac{\mathrm{d}P_\infty}{\mathrm{d}p}. \quad \square$$

3.2.3 A Bound on the Expected Cluster Size.

The next theorem also gives a mean field bound, this time on the susceptibility exponent γ. It was proved by Aizenman and Newman [AN1].

THEOREM 3.8. ([AN1]) *For $p < p_c$,*

$$\chi(p) \ge \frac{1}{c(d)(p_c - p)}$$

where $c(d)$ is the coordination number (i.e. the number of nearest neighbors).

REMARKS. (i) If the exponent γ defined by (1.15) exists, then the theorem implies $\gamma \ge 1$.

(ii) This theorem is analogous to that in the last section in that it also follows from a differential inequality. But here the inequality must be interpreted in a weak sense due to possible discontinuities. Ignoring for the moment these subtleties, the differential inequality is

$$-\frac{\mathrm{d}}{\mathrm{d}p}\frac{1}{\chi(p)} \le c(d).$$

PROOF OF THEOREM 3.8. We start from the relation

$$\chi(p) = \sum_{x\in\mathbb{L}_d} \tau(0,x;p).$$

As in the proof of Theorem 3.5, we would like to differentiate and use Russo's formula. However, here again the event in question depends in principle on an infinite number of bonds. In the proof of Theorem 3.5, we circumvented this difficulty by using the right derivative inequality form of Russo's formula (3.13). Here, however, the inequality goes in the wrong direction, so we instead define finite-volume analogues of all the appropriate quantities and then take the volume to infinity. Let B_n be a box of side n centered at the origin. We define:

$$\begin{aligned}
\mathcal{C}^n(x) &= \{y\in B_n : x\leftrightarrow y \text{ by a path contained in } B_n\},\\
T^n(0,x) &= \{0\in \mathcal{C}^n(x)\},\\
\tau^n(0,x;p) &= P_p(T^n(0,x)),\\
\chi^n(p;x) &= E_p(\mathcal{C}^n(x)),\\
\text{and } \hat{\chi}^n(p) &= \max_{x\in B_n}\chi^n(p;x).
\end{aligned}$$

Since $\chi(p)\ge\chi^n(p;x)$ for every $x\in B_n$, we have

$$\chi(p) \ge \hat{\chi}^n(p),$$

and therefore

$$\chi(p) \ge \limsup_{n\to\infty}\hat{\chi}^n(p).$$

Furthermore,

$$\hat{\chi}^n(p) \ge \chi^n(p;0) = \sum_{x\in B_n}\tau^n(0,x;p).$$

Since $T^n(0,x)$ increases to $T(0,x)$ as n tends to ∞, we have

$$\tau^n(0,x;p)\to\tau(0,x;p).$$

Therefore, by monotone convergence,

$$\liminf_{n\to\infty}\hat{\chi}^n(p) \ge \lim_{n\to\infty}\sum_{x\in B_n}\tau^n(0,x;p) = \sum_{x\in\mathbb{L}_d}\tau(0,x;p) = \chi(p).$$

Combining these inequalities gives

$$\lim_{n\to\infty}\hat{\chi}^n(p) = \chi(p)$$

(even in the case $\chi(p) = \infty$).

Differentiating the expression

$$\chi^n(p;x) = \sum_{z \in B_n} \tau^n(x,z;p) = \sum_{z \in B_n} P_p(T^n(x,z))$$

gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\chi^n(p;x) &= \sum_{z \in B_n} \frac{\mathrm{d}}{\mathrm{d}p} P_p(T^n(x,z)) \\ &= \sum_{z \in B_n} \sum_{b \in \tilde{B}_n} P_p(b \in \delta T^n(x,z))\end{aligned}$$

by Russo's formula, where $\tilde{B}_n$ are the bonds in B_n. Suppose that b is pivotal for $T^n(x,z)$ and let y_b and y_b' denote the endpoints of b. We may write

$$\{b \in \delta T^n(x,z)\} = T^n(x,y_b) \circ T^n(y_b',z) \ \cup \ T^n(x,y_b') \circ T^n(y_b,z).$$

Hence, by the BK inequality,

$$P_p(b \in \delta T^n(x,z)) \ \le \ \tau^n(x,y_b;p)\,\tau^n(y_b',z;p) \ + \ \tau^n(x,y_b';p)\,\tau^n(y_b,z;p).$$

Summing on $z \in B_n$, we obtain

$$\sum_{z \in B_n} P_p(b \in \delta T^n(x,z)) \ \le \ \tau^n(x,y_b;p)\,\chi^n(p;y_b') \ + \ \tau^n(x,y_b';p)\,\chi^n(p;y_b).$$

Therefore,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\chi^n(p;x) &\le \sum_{b=(y_b,y_{b'}) \in \tilde{B}_n} \left(\tau^n(x,y_b;p)\,\chi^n(p;y_b') + \tau^n(x,y_b';p)\,\chi^n(p;y_b)\right) \\ &\le \ \hat{\chi}^n(p) \sum_{b \in \tilde{B}_n} \left(\tau^n(x,y_b;p) + \tau^n(x,y_b';p)\right) \\ &= \ c(d)\,\hat{\chi}^n(p)\,\chi^n(p;x) \\ &\le c(d)\,\hat{\chi}^n(p)^2\,, \end{aligned} \tag{3.18}$$

where $c(d)$ is the number of nearest neighbors and where the third step uses the fact that each site is summed over $\frac{1}{2}c(d)$ times.

Since we are working in a finite box, $\chi^n(p;x)$ is a polynomial, and therefore the maximum over x is differentiable, except perhaps at finitely many p. Furthermore, the derivative of the maximum is less than or equal to the maximum of the derivative. Thus

$$\frac{\mathrm{d}}{\mathrm{d}p}\hat{\chi}^n(p) \ \le \ \max_{x \in B_n} \frac{\mathrm{d}}{\mathrm{d}p}\chi^n(p;x) \ \le \ c(d)\hat{\chi}^n(p)^2$$

by (3.18). Alternatively,

$$-\frac{\mathrm{d}}{\mathrm{d}p}\frac{1}{\hat{\chi}^n(p)} \leq c(d).$$

Integrating from p to p_c, we have

$$\frac{1}{\hat{\chi}^n(p)} - \frac{1}{\hat{\chi}^n(p_c)} \leq c(d)(p_c - p).$$

Finally we can let n tend to ∞. Since $\chi(p_c) = \infty$, we obtain

$$\chi(p) \geq \frac{1}{c(d)(p_c - p)}. \quad \square$$

3.2.4 The Upper Critical Dimension and Saturation of the Mean-Field Bounds.

As we mentioned earlier, when the dimension gets high enough, the system acts as if it is infinite-dimensional. This may be more familiar in the case of random walks. It is well known that two random walks will avoid each other in $d > 4$, but will intersect in $d < 4$. The dimension $d_c = 4$ is the upper critical dimension for random walks. A similar property holds for the Ising magnet ([A], [Fr]). The Ising criterion for mean-field behavior is that the so-called "bubble diagram" satisfies

$$B(\beta) \equiv \sum_x s^2(0, x) < \infty \quad \text{at } \beta = \beta_c,$$

where $s(0, x)$ is the truncated pair correlation (i.e., the Ising analogue of $\tau^{\text{fin}}(0, x)$). Again, this occurs beyond the critical dimension $d_c = 4$.

For percolation, Aizenman and Newman [AN1] showed that if the "triangle diagram" satisfies

$$\nabla(p) := \sum_{x,y} \tau(0, x; p)\, \tau(x, y; p)\, \tau(y, x; p) < \infty \quad \text{at } p = p_c,$$

then β, γ, and δ exist and assume their mean-field values. The presence of three factors of τ reflects the (as yet, unproved) fact that $d_c = 6$ for percolation. Using the lace expansion—a technique developed by Brydges and Spencer [BS] to study self avoiding walks— Hara and Slade [HS] showed that indeed $\nabla(p) < \infty$ if d is sufficiently large. See [MS] for a pedagogical discussion of the lace expansion for self-avoiding walks.

CHAPTER IV

TWO FUNDAMENTAL QUESTIONS

In this chapter, we review two of the major achievements of percolation theory of the past decade: the proofs of $\pi_c = p_c$ and of uniqueness of the infinite cluster.

4.1 Absence of an Intermediate Phase

In this subsection, we discuss the proof of the relation

$$\pi_{\mathrm{c}} = p_{\mathrm{c}}$$

for regular lattices in all dimensions. (Recall that the critical point π_{c} is defined in (1.6).) If this were not the case, then there would be an intermediate phase in which $\chi(p) = \infty$ but $P_\infty(p) = 0$; that is, the cluster of the origin would have infinite expected size, but there would almost surely be no infinite cluster. In that sense, every point of such a phase would be similar to the critical point. It has, in fact, been shown that there is an intermediate phase for percolation on a wedge-like lattice that opens logarithmically slowly [CC2]. (See Figure 1.3.) This counterexample helped to established that translation invariance is a necessary condition for the absence of an intermediate phase.

In 1980, Kesten [K1] showed $\pi_{\mathrm{c}} = p_{\mathrm{c}}$ $(= \frac{1}{2})$ for bond percolation on $\mathbb{B}_2$. For higher dimensions, $\pi_{\mathrm{c}} = p_{\mathrm{c}}$ was finally proved six years later independently by Menshikov, Molchanov, and Siderenko [MMS] and Aizenman and Barsky [AB].

Theorem 4.1. ([MMS], [AB]) *For translation invariant lattices,*

$$\pi_{\mathrm{c}} = p_{\mathrm{c}}.$$

Both the [MMS] and [AB] proofs are related to the [CC3] result

$$P_\infty(p) \ \geq \ p - p_{\mathrm{c}} \tag{4.1}$$

which was proved in the previous chapter (Corollary 3.6). Indeed, a bound of this form is the key ingredient in the [MMS] proof. Aizenman and Barsky consider a generalization of the percolation model with an additional parameter which is comparable to an external field in a spin model. The key estimate in their proof is a differential inequality which is the generalization of the inequality in Theorem 3.5 to a nonzero external field. Before reviewing (part of) the Aizenman and Barsky proof, let us discuss percolation in an external field.

As we will see when we discuss the Potts models, percolation is very much like a spin model, with p related to the inverse temperature β by $p = 1 - e^{-\beta}$. Spin models have an additional parameter, h, called an *external field*. This parameter tends to order the system—when $h \neq 0$, the analogue of P_∞ is alway non-zero. The question of long-range order is the question of whether this analogue of P_∞ remains nonzero as $h \searrow 0$. The analogue of P_∞ is the magnetization, denoted by $M(\beta, h)$. The susceptibility $\chi(\beta, h)$ is the derivative of M with respect to the field h:

$$\chi(\beta, h) = \frac{\partial M(\beta, h)}{\partial h},$$

so that

$$\chi(\beta) = \lim_{h \searrow 0} \frac{\partial M(\beta, h)}{\partial h}$$

is the analogue of $\chi(p)$.

Thus, for percolation, we introduce a new parameter γ (which we think of as $1 - e^{-h}$) and try to define $P_\infty(p, \gamma)$ and $\chi(p, \gamma)$ so that

$$P_\infty(p) = \lim_{\gamma \searrow 0} P_\infty(p, \gamma)$$

and

$$\chi(p) = \lim_{\gamma \searrow 0} \chi(p, \gamma) = \lim_{\gamma \searrow 0} \frac{\partial P_\infty(p, \gamma)}{\partial \gamma}.$$

Using ideas that went back to R. Griffiths [Grif], this was accomplished as follows.

Introduce a new vertex g called a *ghost vertex* and join each site $x \in \mathbb{L}_d$ to g with probability γ. Now the set of realizations is

$$\tilde{\Omega} = \{0, 1\}^{\mathbb{B}_d} \times \{0, 1\}^{\mathbb{L}_d}$$

where the first factor, as usual, represents the bond configurations on $\mathbb{B}_d$, and the second represents the configurations of sites connected to g. Just as each $\omega \in \Omega$ corresponds to a set of bonds $S(\omega) \subset \mathbb{B}_d$ of occupied bonds, each $\omega \in \tilde{\Omega}$ corresponds both to a set $S(\omega) \subset \mathbb{B}_d$ of bonds and a set $\mathcal{G}(\omega) \subset \mathbb{L}_d$ of sites, which we will call "green sites", that connect directly to the ghost site. Notice that, in general, $S(\omega)$ contains many components, but $\mathcal{G}(\omega)$ has only one. We may think of g as an "additional point at infinity" in the sense that all sites connected to g are connected to each other.

The analogue of the infinite cluster density is

$$P_\infty(p, \gamma) = P_{p,\gamma}(\mathcal{C}(0) \cap \mathcal{G} \neq \emptyset). \tag{4.2}$$

As γ decreases to 0, the set $\mathcal{G}$ becomes more and more sparse, so that the cluster of the origin, $\mathcal{C}$, must become larger and larger if it is to have a nontrivial intersection with $\mathcal{G}$. This suggests that

$$\lim_{\gamma \searrow 0} P_\infty(p, \gamma) = P_p(|\mathcal{C}| = \infty) = P_\infty(p). \tag{4.3}$$

To prove this, simply write

$$\begin{aligned}P_\infty(p,\gamma) = 1 - &\sum_{n=1}^{\infty} P_{p,\gamma}(\mathcal{C}\cap\mathcal{G}=\emptyset \mid |\mathcal{C}|=n)\, P_p(|\mathcal{C}|=n)\\ &- P_{p,\gamma}(\mathcal{C}\cap\mathcal{G}=\emptyset \mid |\mathcal{C}|=\infty)\, P_p(|\mathcal{C}|=\infty)\\ = 1 - &\sum_{n=1}^{\infty}(1-\gamma)^n\, P_p(|\mathcal{C}|=n),\end{aligned}$$

since $P_{p,\gamma}(\mathcal{C}\cap\mathcal{G}=\emptyset \mid |\mathcal{C}|=\infty)=0$.

Next, let us define an analogue of $\chi(p)$. We make the obvious choice:

$$\chi(p,\gamma) = E_{p,\gamma}(|\mathcal{C}|,\ \mathcal{C}\cap\mathcal{G}=\emptyset). \tag{4.4}$$

Again, it is clear that, for $p<p_c$,

$$\lim_{\gamma\searrow 0}\chi(p,\gamma) = \chi^{\mathrm{fin}}(p) = \chi(p). \tag{4.5}$$

Indeed, we have

$$\chi(p,\gamma) = \sum_{n=1}^{\infty} n(1-\gamma)^n\, P_p(|\mathcal{C}|=n),$$

which is enough to see (4.5).

Recall that the point here is to show that (after taking γ to zero) $P_\infty(p)$ and $\chi(p)$ have singularities at the same value of p. First, let us show the $P_\infty(p,\gamma)$ and $\chi(p,\gamma)$ are related here as they are related in spin models. To this end, note that $P_\infty(p,\gamma)$ is a power series in $1-\gamma$ with radius of convergence at least 1. So, we can differentiate with respect to γ to get

$$\begin{aligned}\frac{\partial P_\infty(p,\gamma)}{\partial\gamma} &= \sum_{n=1}^{\infty} n(1-\gamma)^{n-1} P_p(|\mathcal{C}|=n)\\ &= \frac{1}{1-\gamma}\,\chi(p,\gamma).\end{aligned}$$

That is, for $0\le\gamma\le 1$,

$$\chi(p,\gamma) = (1-\gamma)\,\frac{\partial P_\infty(p,\gamma)}{\partial\gamma}. \tag{4.6}$$

This, up to a harmless factor of $1-\gamma$, is what we wanted. It is useful in proving the following two differential inequalities:

PROPOSITION 4.2. ([AB]) *If $\gamma\in(0,1)$, $p\in(0,1)$, then*

$$(1-p)\frac{\partial P_\infty(p,\gamma)}{\partial p} \le c(1-\gamma)P_\infty(p,\gamma)\frac{\partial P_\infty(p,\gamma)}{\partial\gamma} \tag{4.7}$$

and

$$P_\infty(p,\gamma) \leq \gamma\frac{\partial P_\infty(p,\gamma)}{\partial\gamma} + P_\infty^2(p,\gamma) + pP_\infty(p,\gamma)\frac{\partial P_\infty(p,\gamma)}{\partial p} \tag{4.8}$$

where $c = c(d)$ is the coordination number of lattice.

Note that the second inequality is the analogue of Theorem 3.5 in an external field. We will discuss the proofs of the two inequalities shortly. Before doing that, however, let us very briefly sketch the proof of Theorem 4.1 given Proposition 4.2.

Brief Sketch of Proof of Theorem 4.1 Given Proposition 4.2. The proof has two main steps:
Step 1: In the first step, we substitute the first inequality (4.7) into the second (4.8) to get a differential inequality purely in terms of P_∞ and $\frac{\partial P_\infty}{\partial\gamma}$. Then standard manipulations—changing variables, integrating, etc.—lead to the following:

Proposition 4.3. ([AB]) *If p is such that $\chi^{\text{fin}}(p) = \infty$, then there is a constant $k = k(p) > 0$ such that*

$$P_\infty(p,\gamma) \geq k\gamma^{1/2}$$

for small $\gamma > 0$.

Remark. It turns out that the exponent δ, defined in (1.19) in terms of the decay of the critical finite cluster distribution, may alternatively be defined by

$$P_\infty(p,\gamma) \approx \gamma^{1/\delta} \qquad\qquad \text{as } \gamma \searrow 0$$

so that Proposition 4.3 gives the mean field bound $\delta \geq 2$.

Step 2: In the second step, we integrate the principal inequality (4.8) over p, using the control from Proposition 4.3 to take the limit $\gamma \searrow 0$. The result is:

Lemma 4.4. ([AB]) *If p is such that $\chi^{\text{fin}}(p) = \infty$, then either*
(a) *$P_\infty(p) > 0$, or*
(b) *$P_\infty(p) = 0$ and, for all $p' \geq p$,*

$$P_\infty(p') \geq \frac{1}{2p'}(p'-p)\,. \tag{4.9}$$

Notice that if we apply this lemma at p_c, then case (a) is just $P_\infty(p_c) > 0$, which we cannot rule out, and, up to a factor of 1/2, case (b) is just Corollary 3.6. We lose the factor of 1/2 in controlling the limit $\gamma \searrow 0$.

Now the rest of the proof is easy. The lemma says that if $\chi^{\text{fin}}(p) = \infty$ and (a) holds, then $p \geq p_c$, while if $\chi^{\text{fin}}(p) = \infty$ and (b) holds, then for all $p' \geq p$, $P_\infty(p') \geq \frac{1}{2p'}(p'-p)$, which also implies $p \geq p_c$. In other words, $\chi^{\text{fin}}(p) = \infty$ implies $p \geq p_c$.

Now suppose $p \in (\pi_c, p_c)$. Then $\chi(p) = \infty$ since $p > \pi_c$, but $\chi(p) = \chi^{\text{fin}}(p)$ since $p < p_c$, so that $\chi^{\text{fin}}(p) = \infty$. But then, by the argument in the above paragraph, $p \geq p_c$, a contradiction. $\square$

From the above sketch, it should be clear that the heart of the proof of Theorem 4.1 is the proof of the inequalities themselves. As in the proof of Theorem 3.8, this proof is complicated by the need to restrict to finite volumes and then to establish convergence of the relevant quantities and their derivatives. Since we have already shown how to deal with problems of this type in Chapter 3, will not concern ourselves here with the infinite-volume limit and will instead simply show some of the finite-volume derivations. For a complete proof, refer to [AB] or [Grim2].

Given a positive integer N, we let Λ_N be the box centered at the origin of size $2N \times 2N$ with periodic boundary conditions, i.e. a $2N \times 2N$ torus. We define

$$\mathcal{G}_N = \mathcal{G} \cap \Lambda_N,$$

$$\mathcal{C}_N(x) = \mathcal{C}(x) \cap \Lambda_N,$$
$$\mathcal{C}_N = \mathcal{C}_N(0),$$
$$A_N(\Gamma) = \{\mathcal{C}_N \cap \Gamma \neq \emptyset\},$$
$$\chi^N(p,\gamma) = E_{p,\gamma}(|\mathcal{C}_N|,\ \mathcal{C}_N \cap \mathcal{G}_N = \emptyset)$$

and

$$\begin{aligned} P^N_\infty(p,\gamma) &= P_{p,\gamma}(\mathcal{C}_N \cap \mathcal{G}_N \neq \emptyset) \\ &= P_{p,\gamma}(A_N(\mathcal{G}_N)). \end{aligned}$$

Sketch of Proof of the First Inequality of Proposition 4.2. We begin by decomposing $\partial P^N_\infty(p,\gamma)/\partial p$ with respect to G_N. We have

$$\begin{aligned} \frac{\partial}{\partial p} P^N_\infty(p,\gamma) &= \frac{\partial}{\partial p} \sum_\Gamma P_{p,\gamma}(\mathcal{G}_N = \Gamma)\, P_{p,\gamma}(A_N(\Gamma) | \mathcal{G}_N = \Gamma) \\ &= \sum_\Gamma P_{p,\gamma}(\mathcal{G}_N = \Gamma)\, \frac{\mathrm{d}}{\mathrm{d}p} P_p(A_N(\Gamma)), \end{aligned} \tag{4.10}$$

where we have used that
(i) the sum runs over the finite number of possible sets $\mathcal{G}_N$,
(ii) $P_{p,\gamma}(\mathcal{G}_N = \Gamma)$ does not depend on p,
(iii) $P_{p,\gamma}(A_N(\Gamma))$ does not depend on γ, and
(iv) $A_N(\Gamma)$ is independent of the event $\{\mathcal{G}_N = \Gamma\}$.

The key step is to rewrite the second factor in the sum using Russo's formula (Theorem 3.4):

$$\frac{\mathrm{d}}{\mathrm{d}p} P_p(A_N(\Gamma)) = \sum_{b \in \Lambda_N} P_p(b \in \delta A_N(\Gamma)).$$

To evaluate this, let b be a bond in Λ_N and suppose for the moment that b is pivotal for $A_N(\Gamma)$. Further suppose that b is vacant. Then one vertex of b must connect to the origin, but not to Γ, while the other vertex of b must connect to Γ, but not to the origin. Conversely, if the vertices of b have those properties, then b is vacant and is pivotal for $A_N(\Gamma)$. That is, letting v_{-1} and v_1 denote the endpoints of b,

$$P_p(b \text{ is vacant}, \, b \in \delta A_N(\Gamma)) = \sum_{i \in \{-1,1\}} P_p(v_i \in \mathcal{C}_N, \ \mathcal{C}_N \cap \Gamma = \emptyset, \ \mathcal{C}_N(v_{-i}) \cap \Gamma \neq \emptyset).$$

Recalling that $\{b \in \delta A_N(\Gamma)\}$ is independent of the status of bond b,

$$(1-p)\frac{\mathrm{d}}{\mathrm{d}p} P_p(A_N(\Gamma)) = \sum_{(x,y)} P_p(x \in \mathcal{C}_N, \ \mathcal{C}_N \cap \Gamma = \emptyset, \ \mathcal{C}_N(y) \cap \Gamma \neq \emptyset), \tag{4.11}$$

where the sum is over ordered pairs (x,y) that form a bond in Λ_N. It follows from (4.10) and (4.11) that

$$(1-p)\frac{\partial P_\infty^N}{\partial p} = \sum_{(x,y)} P_{p,\gamma}(x \in \mathcal{C}_N, \ \mathcal{C}_N \cap \mathcal{G}_N = \emptyset, \ \mathcal{C}_N(y) \cap \mathcal{G}_N \neq \emptyset).$$

Next, we make a decomposition with respect to $\mathcal{C}_N$:

$$(1-p)\frac{\partial P_\infty^N}{\partial p} = \sum_{(x,y)} \sum_{\Delta} P_p(\mathcal{C}_N = \Delta) \, P_{p,\gamma}(\mathcal{C}_N \cap \mathcal{G}_N = \emptyset, \ \mathcal{C}_N(y) \cap \mathcal{G}_N \neq \emptyset \,|\, \mathcal{C}_N = \Delta),$$

where the second sum is over connected sets $\Delta \subset \Lambda_N$ such that $0, x \in \Delta$ and $y \notin \Delta$.

Now, conditioned on $\mathcal{C}_N = \Delta$, $\mathcal{C}_N \cap \mathcal{G}_N = \emptyset$ depends only on sites in Δ, whereas $\mathcal{C}_N(y) \cap \mathcal{G}_N \neq \emptyset$ depends only on sites and bonds outside of Δ. Hence, $\mathcal{C}_N \cap \mathcal{G}_N = \emptyset$ and $\mathcal{C}_N(y) \cap \mathcal{G}_N \neq \emptyset$ are independent, conditioned on $\mathcal{C}_N = \Delta$.

Furthermore, since $y \notin \Delta$,

$$\begin{aligned} P_{p,\gamma}(\mathcal{C}_N(y) \cap \mathcal{G}_N \neq \emptyset \,|\, \mathcal{C}_N = \Delta) &= P_{p,\gamma}(\mathcal{C}_N(y) \cap \mathcal{G}_N \neq \emptyset \,|\, \Delta \not\leftrightarrow \Delta^c) \\ &\leq P_{p,\gamma}(\mathcal{C}_N(y) \cap \mathcal{G}_N \neq \emptyset) \\ &= P_\infty^N(p,\gamma), \end{aligned}$$

where the second step uses the FKG inequality and the third step uses translation invariance.

Using this, and recalling that the sum over Δ is constrained to satisfy $0, x \in \Delta$ and $y \notin \Delta$, we have

$$\begin{aligned} (1-p)\frac{\partial P_\infty^N}{\partial p} &\leq \sum_{(x,y)} \sum_{\Delta} P_p(\mathcal{C}_N = \Delta) \, P_{p,\gamma}(\mathcal{C}_N \cap \mathcal{G}_N = \emptyset \mid \mathcal{C}_N = \Delta) \, P_\infty^N(p,\gamma) \\ &= P_\infty^N(p,\gamma) \sum_{(x,y)} P_{p,\gamma}(x \in \mathcal{C}_N, \ y \notin \mathcal{C}_N, \ \mathcal{C}_N \cap \mathcal{G}_N = \emptyset) \\ &\leq P_\infty^N(p,\gamma) \sum_{(x,y)} P_{p,\gamma}(x \in \mathcal{C}_N, \ \mathcal{C}_N \cap \mathcal{G}_N = \emptyset) \\ &\leq c \, P_\infty^N(p,\gamma) \sum_{x \in \Lambda_N} P_{p,\gamma}(x \in \mathcal{C}_N, \ \mathcal{C}_N \cap \mathcal{G}_N = \emptyset) \\ &= c \, P_\infty^N(p,\gamma) \, E_{p,\gamma}(|\mathcal{C}_N|, \ \mathcal{C}_N \cap \mathcal{G}_N = \emptyset) \\ &= c \, P_\infty^N(p,\gamma) \, \chi^N(p,\gamma), \end{aligned}$$

where c is the number of nearest neighbors.

But now recall (4.6):

$$\chi = (1-\gamma)\frac{\partial P_\infty}{\partial \gamma}.$$

The same relationship exists between χ^N and P_∞^N. So,

$$(1-p)\,\frac{\partial P_\infty^N(p,\gamma)}{\partial p} \;\le\; c\,(1-\gamma)\,P_\infty^N(p,\gamma)\,\frac{\partial P_\infty^N(p,\gamma)}{\partial \gamma}.$$

To complete the proof, we would now have to take the limit as N goes to infinity, proving convergence of P_∞^N and its derivatives. □

Sketch of Proof of the Second Inequality of Proposition 4.2. We decompose P_∞^N according to:

$$\begin{aligned}
P_\infty^N(p,\gamma) &= P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| = 1) \;+\; P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| \ge 2)\\
&= \sum_{n=1}^{\infty} P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| = 1 \,|\, |\mathcal{C}_N| = n)\, P_p(|\mathcal{C}_N| = n)\\
&\qquad + \; P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| \ge 2)\\
&= \sum_{n=1}^{\infty} n\gamma(1-\gamma)^{n-1}\, P_p(|\mathcal{C}_N| = n) \;+\; P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| \ge 2)\\
&= \frac{\gamma}{1-\gamma}\,\chi^N(p,\gamma) \;+\; P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| \ge 2)\\
&= \gamma\,\frac{\partial P_\infty^N(p,\gamma)}{\partial \gamma} \;+\; P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| \ge 2),
\end{aligned}$$

where we have again used (4.6) in the last step. So, we have reduced the problem to analyzing $P_{p,\gamma}(|\mathcal{C}_N \cap \mathcal{G}_N| \ge 2)$. This is the hard part of the proof, but it is essentially identical to the proof of Theorem 3.5 except that the connections to infinity there are replaced by connections to $\mathcal{G}_N$ here. With these ideas in mind, the condition $|\mathcal{C}_N \cap \mathcal{G}_N| \ge 2$ relates to being in the backbone or in a dangling end, whereas $|\mathcal{C}_N \cap \mathcal{G}_N| = 1$ relates to being spineless. Once again, the proof ends with limits in N. □

4.2 Uniqueness of the Infinite Cluster

In 1960, Harris [Har] proved that the infinite cluster is unique in two dimensions. Unfortunately, his proof was not generalizable as it relied heavily on the fact that the dual of a bond is a bond in two dimensions (i.e., in $d = 2$, codimension 1 objects are of dimension 1). In 1981, Newman and Schulman [NS1] gave a soft, but beautiful argument, which we will discuss below, showing that, under very general conditions, the number of clusters $N = N(\omega)$ is either 0, 1, or ∞ with probability 1. This left a big open problem: prove that $N = 1$ for $p > p_c$. Of course, this can not be true on all graphs—for example, $N = \infty$ on a homogeneous tree. Indeed, for some time the necessary conditions were not clear.

After seven years, a burst of activity ensued. In 1987, Aizenman, Kesten, and Newman [AKN] proved uniqueness for both nearest-neighbor and long-range independent percolation models on $\mathbb{Z}^d$. Their proof was simplified the following year by Gandolfi, Grimmett and Russo [GGR]. Then, in 1989, Gandolfi [Ga] proved uniqueness for stationary Gibbs states. Meanwhile, in 1988, Gandolfi, Keane, and Russo [GKR] extended the two-dimensional Harris [Har] uniqueness result to models which obeyed the FKG condition and were ergodic under translations in each lattice direction. Finally, in 1989, Burton and Keane [BuK] gave a simple and beautiful proof of uniqueness for nearest-neighbor model on $\mathbb{Z}^d$ under the conditions of the earlier [NS1] result—namely, stationarity and "finite energy". In 1992, the Burton and Keane result was generalized to long-range models on $\mathbb{Z}^d$ or $\mathbb{Z}^d \times \mathbb{N}$ by Gandolfi, Keane, and Newman [GKN]. Here, we will discuss the [NS1] and [BuK] proofs. First, however, let us review the definitions of the required conditions.

Throughout this subsection, we will restrict attention to a finite-dimensional lattice $\mathbb{L}^d$, and for convenience, we will consider configurations of sites rather than configurations of bonds.

Stationarity.

Let $x \in \mathbb{L}_d$. Define a shift operator T_x acting on configurations $\omega \in \Omega = \{0,1\}^{\mathbb{L}_d}$, events $A \in \mathcal{F}$, measures P, and random variables X on Ω by

$$(T_x\omega)_y = \omega_{y-x},$$

$$T_xA = \{T_x\omega : \omega \in A\},$$

$$(T_xP)(A) = P(T_{-x}A),$$

and

$$(T_xX)(\omega) = X(T_{-x}\omega).$$

Definition: An event A is translation invariant if $T_xA = A$ for all $x \in \mathbb{L}_d$. A measure P is *stationary* if $T_xP = P$ for all $x \in \mathbb{L}_d$. P is *(translation) ergodic* if every invariant event A is P-trivial, i.e. $P(A) = 0$ or $P(A) = 1$.

If P is stationary but not ergodic, then P can be decomposed into a convex combination of *ergodic components*. (In physics, these components would be identified with "pure phases".) So, without loss of generality, we can take our stationary P to be ergodic.

We complete our review of stationarity and ergodicity with the following well known theorem.

Theorem 4.5. (Birkhoff Ergodic Theorem) *Let P be a (translation) ergodic measure on some Ω and let $X \in L^1(\Omega, P)$. Then*

$$\frac{1}{|\Lambda|}\sum_{j\in\Lambda} T_jX \to E(X) \quad \textit{a.s. and in } L^1$$

as $\Lambda \nearrow \mathbb{L}_d$.

The Finite-Energy Condition.

DEFINITION: P is a *finite-energy measure* if $P(\cdot|\omega(x)=1)$ and $P(\cdot|\omega(x)=0)$ are equivalent for all $x \in \mathbb{L}_d$; that is, P has finite energy if, for every $x \in \mathbb{L}_d$ and $A \in \mathcal{F}_{\mathbb{L}_d \setminus \{x\}}$,

$$P(A|\omega(x)=0)=0 \iff P(A|\omega(x)=1)=0.$$

We hasten to point out that independent site (or bond) percolation has finite energy. Another finite-energy measure comes up in later chapters when we discuss the Ising magnet and, more generally, the Potts model in both the standard spin representation and the FK representation. The property described here is called "finite energy" since, for any Gibbs measure, the Radon-Nikodym derivative of one conditional measure with respect to the other is just the exponential of the energy difference. If the energy difference is finite, then the Radon-Nikodym derivative does not vanish in a finite volume and hence the measures are equivalent. The following proposition gives another interpretation of the finite-energy condition.

PROPOSITION 4.6. *Let $S \subset \mathbb{L}_d$ be a finite set and let $\phi_S \in \{0,1\}^S$ be a configuration on S. For each configuration $\omega \in \Omega$, define its modified version*

$$\phi_S(\omega) = \begin{cases} \phi_S(x) & \text{if } x \in S \\ \omega(x) & \text{if } x \notin S \end{cases}$$

and, for each event $A \in \mathcal{F}$, define its modified version

$$\phi_S(A) = \{\phi_S(\omega) : \omega \in A\}. \tag{4.12}$$

Then P is a finite-energy measure if and only if for all $A \in \mathcal{F}$ and all finite $S \subset \mathbb{L}_d$

$$P(A) > 0 \implies P(\phi_S(A)) > 0.$$

The proof of this proposition is an induction proof left to the reader. For future reference, notice it follows from (4.12) that, for an arbitrary event A,

$$\phi_S(A) \;=\; \{\omega : \exists\, \omega' \in A \quad \omega = \phi_S(\omega')\}, \tag{4.13}$$

and therefore for events which satisfy $\phi_S(A) = A$,

$$\phi_S(A) \;=\; \{\omega \in A : \omega = \phi_S(\omega)\}\,. \tag{4.14}$$

We end this brief discussion of stationarity and the finite-energy condition by pointing out that, when decomposing a stationary finite energy measure into ergodic components, each component still retains the finite-energy property. The following proposition was implicitly assumed, but not stated or proved in [BuK]; it was proved in a general context by Gandolfi, Keane and Newman [GKN]. The proof below is due to Borgs and Chayes.

PROPOSITION 4.7. *If P is a finite-energy measure, then all elements in the ergodic decomposition of P also have finite energy.*

PROOF. Here we will restrict attention to measures P for which the ergodic decomposition is of the form of a finite convex combination

$$P = \sum_{\alpha} \lambda_\alpha P_\alpha ,$$

where each ergodic component P_α is a conditional measure with respect to a nontrivial event A_α in the tail field, i.e. an event A_α in the tail field such that $0 < P(A_\alpha) < 1$. The reader is referred to [GKN] for a proof of the most general case.

Let P_α be an ergodic component of P and let A_α be the corresponding nontrivial event in the tail field, so that

$$P_\alpha = P(\,\cdot \mid A_\alpha). \tag{4.15}$$

By Proposition 4.6, it suffices to show that if ϕ is any finite modification of the form (4.12) and B is any event, then

$$P_\alpha(B) > 0 \implies P_\alpha(\phi(B)) > 0 ,$$

which, by (4.15) and the fact that $0 < P(A_\alpha) < 1$, is equivalent to

$$P(B \cap A_\alpha) > 0 \implies P(\phi(B) \cap A_\alpha) > 0 . \tag{4.16}$$

Now, for any two events A and B, it follows from (4.13) and (4.14) that

$$\phi(B \cap A) \subset \phi(B) \cap \phi(A) .$$

Here $\phi(A_\alpha) \subset A_\alpha$ since A_α is in the tail field. So

$$\phi(B \cap A_\alpha) \subset \phi(B) \cap \phi(A_\alpha) \subset \phi(B) \cap A_\alpha .$$

Using this and the fact that P itself has finite energy, we therefore obtain

$$P(B \cap A_\alpha) > 0 \implies P(\phi(B \cap A_\alpha)) > 0 \implies P(\phi(B) \cap A_\alpha) > 0 ,$$

which is the desired result (4.16). □

The aforementioned results in [NS1] and [BuK] can now be stated explicitly and proved. For the following, our configuration space is $\Omega = \{0,1\}^{\mathbb{L}_d}$, with $\mathbb{L}_d$ a finite-dimensional site lattice, P a measure on Ω, and $N = N(\omega)$ the number of distinct infinite clusters in $\omega \in \Omega$.

THEOREM 4.8. ([NS1]) *If P is stationary and has finite energy, then each ergodic component of P has $N = 0$, 1, or ∞ almost surely.*

PROOF OF THEOREM 4.8. This proof is just a straightforward application of the assumptions. By ergodic decomposition and Proposition 4.7, we can assume without loss of generality that P is ergodic. Then, ergodicity implies that the event $\{N = k\}$ has probability 0 or 1. Let k_0 be the value with $P(N = k_0) = 1$. Assume $1 < k_0 < \infty$. Since $k_0 > 1$, we can find a (large) integer M such that, with positive probability, at least two infinite clusters intersect a box of side M about the origin. Now let us locally modify each configuration by occupying every site in the box. By finite energy in the form of Proposition 4.6, this modified event has positive probability. But, since $k_0 < \infty$, our modified configurations must now have fewer than k_0 infinite clusters, since we have combined at least two of our clusters into one. Since this occurs with positive probability, we have contradicted the assumption that $P(N = k_0) = 1$. □

THEOREM 4.9. ([BuK]) *If P is stationary and has finite energy, then each ergodic component of P has $N = 0$ or 1 almost surely.*

PROOF OF THEOREM 4.9. Again, by Proposition 4.7, we can assume ergodicity. By Theorem 4.8, we need only prove that $P(N = \infty) = 0$. So, we suppose that $N = \infty$ a.s. Now, given a site $x \in \mathbb{L}_d$, x is said to be an *encounter point* in configuration ω if the following hold:

(1) $|\mathcal{C}(x; \omega)| = \infty$ (i.e. x belongs to an infinite cluster of ω)
(2) the set $\mathcal{C}(x; \omega) \setminus \{x\}$ has no finite components and exactly three infinite components.

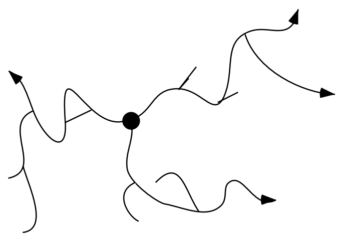


FIGURE 4.1. *An encounter point with arrows representing paths to infinity.*

Just as we used finite energy in the proof of Theorem 4.8, we can use it here to see that the origin is an encounter point with positive probability, say 2ε. Then, by the ergodic theorem, we can almost surely find a large enough rectangle R (depending on ω) so that the number of encounter points in R is at least $\varepsilon|R|$.

However, we will show that the number of encounter points in R must be less than $|\partial R|$. This will give a contradiction since

$$\varepsilon|R| < |\partial R|$$

is false for a large enough rectangle R in finite dimension d.

The heuristics of the remainder of the proof are straightforward: The encounter points of a given cluster are connected by a tree structure; therefore, the number of encounter points in that cluster is bounded by the number of boundary points in the cluster. This idea is formalized as follows:

Let $\mathcal{C} = \mathcal{C}(\omega)$ be an infinite cluster, let

$$Y = \mathcal{C} \cap \partial R,$$

and let $x, x' \in R$ be two encounter points in $\mathcal{C}$. So, $\mathcal{C} \setminus \{x\}$ naturally corresponds to a partition of Y:

$$P = \{P_1, P_2, P_3\}$$

where all the points in P_i are in the same component of $\mathcal{C} \setminus \{x\}$. We have a similar partition using x':

$$Q = \{Q_1, Q_2, Q_3\}.$$

Notice that the indices 1,2, and 3 may be chosen so that

$$P_1 \supset Q_2 \cup Q_3. \tag{4.17}$$

Indeed, this follows immediately by taking P_1 to be the component of $\mathcal{C} \setminus \{x\}$ containing x' and taking Q_1 be the component of $\mathcal{C} \setminus \{x'\}$ containing x. (See Figure 4.2).

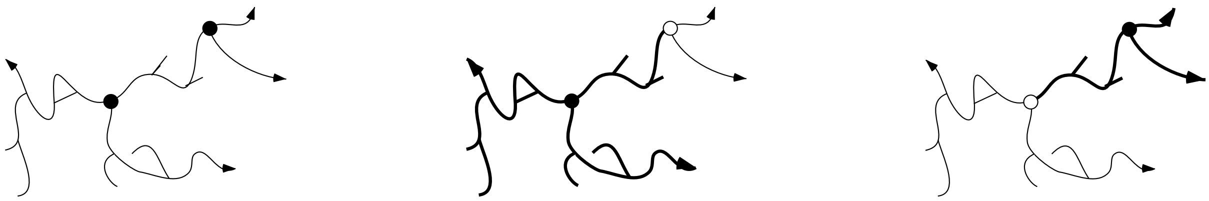


FIGURE 4.2. *An illustration of how to choose the indices of the partitions to satisfy (4.17). The first figure shows a cluster $\mathcal{C}$ with two encounter points. The second figure removes point x and highlights the component that should be labeled P_1. The third figure removes point x' and highlights the component that should be labeled Q_1.*

We will say that any two partitions $P = \{P_1, P_2, P_3\}$ and $Q = \{Q_1, Q_2, Q_3\}$ are *compatible* if there is an ordering of the indices such that (4.17) holds, and that a collection $\mathcal{P}$ of partitions is compatible if each pair $P, Q \in \mathcal{P}$ is compatible. The proof is now a routine consequence of the following lemma:

LEMMA 4.10. *If $\mathcal{P}$ is a compatible collection of partitions of Y, then*

$$|\mathcal{P}| \leq |Y| - 2.$$

Indeed, by summing over all infinite clusters $\mathcal{C}$, we have that the number of encounter points in R is bounded above by $|\partial R|$, a contradiction. $\square$

We will not bother to prove the lemma here; the proof is not difficult and can be quickly read in [BuK]. The key observation, seen in Figure 4.2, is the finite-dimensional intuition discussed above: namely, the only way to get compatible partitions is to have a tree structure. It follows that this proof does not work on a tree, since the inequality $\varepsilon R < |\partial R|$ does not lead to a contradiction there. In fact, on a tree, there are infinitely many infinite clusters.

CHAPTER V

FINITE-SIZE SCALING AND THE INCIPIENT INFINITE CLUSTER

In this chapter, we return to the study of the critical regime and review the new results of Borgs, Chayes, Kesten and Spencer [BCKS] on finite-size scaling and the incipient infinite cluster. We develop in some detail the ideas of length scales and finite-size scaling introduced in Chapter 2.

We consider percolation in a finite box. As a function of the size of this box, we determine the scaling window about p_c in which the system behaves critically. Here criticality is characterized by the behavior of the distribution of sizes of the largest clusters in the box. We show how these clusters can be identified with the so-called incipient infinite cluster—the cluster of infinite expected size which appears at p_c. It turns out that all of these results can be established axiomatically from hypotheses which are mathematical expressions of the purported scaling behavior in critical percolation. Moreover, these hypotheses can be explicitly verified in two dimensions. Here we will omit almost all details of the proofs, although we will discuss the motivation, the hypotheses and certain results at some length. The only proof we will explicitly review is a proof of the scale invariance of the cluster size distribution, given the axioms. The reader is referred to [BCKS] for more details and for related results which are not included here.

5.1 The Motivation

The motivation for the [BCKS] work was threefold, and is perhaps as interesting as the results.

The Random Graph Model.

The initial motivation was to obtain an analogue of recent results on the so-called random graph model of Erdös and Renyi ([ER1], [ER2]). The random graph model is just a mean-field percolation model on N sites in which each site is connected to each other site independently with uniform probability $p(N)$, i.e. it is just percolation on the complete graph. It turns out that the model has nontrivial behavior if $p(N)$ scales like

$$p(N) \approx \frac{c}{N}$$

with $c = \Theta(1)$. Here, as usual, $f = \Theta(N^\alpha)$ means that there are nonzero, finite constants c_1 and c_2 such that $c_1 N^\alpha \leq f \leq c_2 N^\alpha$. Clearly, if c grows more rapidly with N, the model will just collapse into one cluster.

Let $W^{(i)}$ denote the random variable representing the size of the i^{th} largest cluster in the system. Erdös and Renyi showed that the model has a phase transition at $c = 1$ characterized by the behavior of $W^{(1)}$. For $c < 1$,

$$W^{(1)} = \Theta(\log N)$$

with probability one, whereas for $c > 1$,

$$W^{(1)} = \Theta(N),$$

and indeed

$$\frac{W^{(1)}}{N} \to \delta > 0$$

with probability one. On the other hand, for all $c \neq 1$, $W^{(2)} = \Theta(\log N)$. The $\Theta(N)$ cluster for $c > 1$ is clearly the analogue of the infinite cluster in percolation on finite-dimensional graphs; here it is called the *giant component*. As we will see, the $\Theta(\log N)$ clusters are analogues of finite clusters in ordinary percolation.

In the past few years, there has been a tremendous amount of work and remarkable progress on the random graph model. Much of this work culminated in the combinatoric tour de force of Janson, Knuth, Luczak and Pittel [JKLP]. Using remarkably detailed calculations, it was shown that the correct parameterization of the critical regime is

$$p(N) = \frac{1}{N} + \frac{\lambda_N}{N^{4/3}},$$

in the sense that if $N \to \infty$ with $\lim\limits_{N\to\infty} |\lambda_N| < \infty$, then

$$W^{(i)} = \Theta(N^{2/3})$$

for all i, and furthermore each $W^{(i)}$ has a nontrivial distribution (i.e. $W^{(i)}/N^{2/3} \nrightarrow$ constant). On the other hand, if $N \to \infty$ with $\lim\limits_{N\to\infty} \lambda_N = -\infty$, then

$$\frac{W^{(2)}}{W^{(1)}} \to 1$$

with probability one, whereas if $N \to \infty$ with $\lim\limits_{N\to\infty} \lambda_N = +\infty$, then

$$\frac{W^{(2)}}{W^{(1)}} \to 0$$

and

$$\frac{W^{(1)}}{N^{2/3}} \to \infty$$

with probability one. The largest component in the regime with $\lambda_N \to +\infty$ is called the *dominant component*. As we will see, it has an analogue in ordinary percolation.

[BCKS] asked whether there is some finite-dimensional analogue of these results. In order to answer this, they considered d-dimensional percolation in a box of linear size n, and hence volume $N = n^d$. In particular, they asked how the size of the largest cluster in the box behaves as a function of n for $p < p_c$, $p = p_c$ and $p > p_c$. Also, they asked whether there is some window $p(n)$ about p_c such that the system has a nontrivial cluster size distribution within the window.

Finite-Size Scaling.

The considerations of the previous paragraph lead us immediately to the question of finite-size scaling (FSS). Phase transitions cannot occur in finite volumes, since all relevant functions are polynomials and thus analytic; nonanalyticities only emerge in the infinite-volume limit. What quantities should we study to see the phase transition emerge as we go to larger and larger volume?

Before the [BCKS] work, this question had been addressed rigorously only in systems with first-order transitions—transitions at which the correlation length and order parameter are discontinuous. Finite-size scaling at first-order phase transitions was studied in some detail by Borgs and collaborators ([BoK], [BKM], [BI1], [BI2]). At first-order transitions, the problem is simpler because the order parameter jumps at the transition point. Thus the analogue of the behavior $W^{(1)} = \Theta(N^{2/3})$ is $W^{(1)} = \Theta(N)$ at p_c. Moreover, this means that the analogue of the window about p_c is also trivial: it scales like $p(N) = p_c + \lambda_N/N$. In this case, it is possible to ask much more delicate questions by looking at the derivative of the order parameter. Then one can ask how the transition point is shifted and what function describes the smoothing of the discontinuity in finite volume. This was done in some detail in the papers cited above.

Returning now to the question of rigorous finite-size scaling at second-order phase transitions, we see that the issue is complicated by the fact that the order parameter vanishes at the critical point. For example, in percolation it is expected that there is no infinite cluster at p_c, with probability one. However, physicists routinely talk about an incipient infinite cluster at p_c. This brings us to our third motivation.

The Incipient Infinite Cluster.

At p_c, there is no infinite cluster with probability one, but the expected size of the cluster of the origin is infinite. Physicists call this object the incipient infinite cluster (IIC).

In the mid 1980's there were two attempts to construct rigorously an object that could be identified as an incipient infinite cluster. Kesten [K3] proposed to look at the conditional measure in which the origin is connected to the boundary of B_n, a $2n \times 2n$ box centered at the origin, by a path of occupied bonds:

$$P_p^n(\cdot) = P_p(\cdot \mid 0 \leftrightarrow \partial B_n).$$

Observe that, at $p = p_c$, as $n \to \infty$, $P_p^n(\cdot)$ becomes mutually singular with respect to the unconditioned measure $P_p(\cdot)$. Nevertheless, Kesten found that

$$\lim_{n\to\infty} P_{p_c}^n(\cdot) = \lim_{p \searrow p_c} P_p(\cdot \mid 0 \leftrightarrow \infty).$$

Moreover, Kesten studied properties of the infinite object so constructed and found that it has a nontrivial fractal dimension which agrees with the fractal dimension of the physicists' incipient infinite cluster.

Another proposal was made by Chayes, Chayes and Durrett [CCD]. They modified the standard measure in a different manner than Kesten, replacing the uniform p by an inhomogeneous $p(x)$ which varies with the distance from the origin:

$$p(x) = p_c + \frac{c}{|x|^{\zeta}}. \tag{5.1}$$

The idea was to enhance the density just enough to obtain a nontrivial infinite object. [CCD] found that when $\zeta = \frac{1}{\nu}$, where ν is the correlation length exponent, the measure $P_{p(x)}$ has some properties reminiscent of the physicists' incipient infinite cluster.

In the work to be discussed here, [BCKS] propose yet a third rigorous incipient cluster—namely the largest cluster in a box. This is, in fact, exactly the definition that numerical physicists use in simulations. Moreover, it will turn out to be closely related to the IICs constructed by Kesten and Chayes, Chayes and Durrett. Like the IIC of [K3], the largest cluster in a box will have a fractal dimension which agrees with that of the physicists' IIC. Also, the [BCKS] proofs rely heavily on technical estimates from the IIC construction of [K3]. More interestingly, the form of the scaling window $p(n)$ for the [BCKS] problem will turn out to be precisely the form (5.1) of the enhanced density used to construct the IIC of [CCD].

5.2 Definitions of Relevant Quantities and Preliminaries

Much of this subsection is taken almost verbatim from the definition section of a prelimary version of [BCKS]. Throughout this chapter, we will consider only nearest-neighbor bond percolation on the hypercubic lattice. In Chapter 1, we defined the relevant notion of occupied clusters in bond percolation. Here we begin by defining the analogous notion for clusters in a finite box $\Lambda \subset \mathbb{Z}^d$. The *connected* component of the occupied cluster of x within Λ is denoted by $\mathcal{C}_\Lambda(x) = \mathcal{C}_\Lambda(x;\omega)$. Notice that $\mathcal{C}_\Lambda(x)$ does not include pieces of the cluster of x in Λ which are connected to x solely through paths in $\Lambda^c = \mathbb{Z}^d \setminus \Lambda$. We will use $\mathcal{C}_\Lambda^{(1)}, \mathcal{C}_\Lambda^{(2)}, \cdots \mathcal{C}_\Lambda^{(k)}$ to denote the occupied clusters in Λ, ordered from largest to smallest size, with lexicographic order between clusters of the same size. $W_\Lambda^{(i)} = |\mathcal{C}_\Lambda^{(i)}|$ denotes the size of the i^{th} largest cluster in Λ. Finally

$$N_\Lambda(s_1, s_2) = |\{i \mid s_1 \leq W_\Lambda^{(i)} \leq s_2\}| \tag{5.2}$$

denotes the number of clusters in Λ with size between s_1 and s_2, and

$$\tilde{N}_\Lambda(s_1, s_2) = |\{i \mid s_1 \leq W_\Lambda^{(i)} \leq s_2,\ C_\Lambda^{(i)} \not\leftrightarrow \partial\Lambda\}| \tag{5.3}$$

is the corresponding number of clusters which do not touch the boundary $\partial\Lambda$ of Λ. Here, as usual, $\partial\Lambda$ is the set of points $x \in \Lambda$ that have distance less or equal 1 from Λ^c.

In Chapter 1, we defined the (point-to-point) connectivity functions $\tau(x, y; p)$ and $\tau^{\mathrm{fin}}(x, y; p)$, cf. equations (1.11) and (1.12). Here we also define the point-to-plane connectivity function

$$\tilde{\pi}_n(p) = P_p(\exists\ x = (n, \cdot)\ ,\ 0 \leftrightarrow x)\,,$$

and the point-to-box connectivity function

$$\pi_n(p) = P_p(0 \leftrightarrow \partial B_n(0))\,,$$

where

$$B_n(x) = \{y \in \mathbb{Z}^d \mid |x - y| \le n\}$$

and $|\cdot|$ denotes the ℓ_∞-norm. Notice that $\pi_n(p)$ and $\tilde{\pi}_n(p)$ are equivalent in the sense that

$$\tilde{\pi}_n(p) \le \pi_n(p) \le 2d\tilde{\pi}_n(p)\,.$$

A quantity which for $p > p_c$ behaves much like $\tau^{\mathrm{fin}}(x, y; p)$ is the covariance:

$$\tau^{\mathrm{cov}}(x, y; p) = \mathrm{Cov}_p(x \leftrightarrow \infty; y \leftrightarrow \infty)\,,$$

where

$$\mathrm{Cov}_p(A, B) = P_p(A \cap B) - P_p(A)P_p(B)\,.$$

Notice that, by translation invariance, the connectivities $\tau(x, y; p)$ and $\tau^{\mathrm{fin}}(x, y; p)$ are related to the susceptibility and the finite-cluster susceptibility, cf. (1.5) and (1.9), by

$$\chi(p) = E_p(|\mathcal{C}(0)|) = \sum_x \tau(0, x; p)\,,$$

and

$$\chi^{\mathrm{fin}}(p) = E_p(|\mathcal{C}(0)|, |\mathcal{C}(0)| < \infty) = \sum_x \tau^{\mathrm{fin}}(0, x; p) = \sum_{s<\infty} sP_s(p)\,.$$

Similarly, we define

$$\chi^{\mathrm{cov}}(p) = \sum_x \tau^{\mathrm{cov}}(0, x; p)\,.$$

Finally, we introduce the quantity

$$s(n) = (2n)^d\, \pi_n(p_c)\,.$$

It will turn out that $s(n)$ represents the size of the largest critical clusters on scale n. In order to describe these, we define the exponent ρ by

$$\pi_n(p_c) \approx n^{-1/\rho}, \tag{5.4}$$

assuming this exponent exists. Then

$$s(n) \approx n^{d_f} \quad \text{where} \quad d_f = d - \frac{1}{\rho}$$

represents the fractal dimension of the incipient infinite cluster.

In Chapter 2, we discussed the notion of length scales and finite-size scaling lengths. We proved the existence of a fundamental correlation length, $\xi(p)$, (Proposition 2.6) and showed that it is non-trivial for $p < p_c$ (Theorem 2.9). It turns out that for all $p \in [0, 1]$, we can define

$$\xi^{-1}(p) = -\lim_{|x|\to\infty} \log \tau^{\mathrm{fin}}(0, x; p), \tag{5.5}$$

with x taken along a coordinate axis, and show that $\xi(p) < \infty$ for $p \neq p_c$. For $p < p_c$, this is just the result mentioned above. For $p > p_c$, this is a result of [CCGKS]. The proof is rather technical and will not be discussed here. However, the heuristics of the generalization of this proof to the Potts model are discussed in some detail in Section 7.7.

In Chapter 2, we also saw that for $p < p_c$, $\xi(p)$ is equivalent to a finite-size scaling correlation length, i.e. a length defined in terms of events that occur on finite sets. In equations (2.22) and (2.23), we defined $L_0(p) = L_0(p; 2, c/e)$ where $c = c(d)$ is the constant needed to satisfy the hypothesis of the rescaling lemma. In this chapter, we instead take $L_0(p) = L_0(p; 3, \epsilon)$, i.e.

$$L_0(p) = \min\{L \geq 1 : R_{\frac{1}{3},3L} \leq \epsilon\} \quad \text{for} \quad p < p_c, \tag{5.6}$$

where $\epsilon = \epsilon(d)$ has been chosen to satisfy the conditions of the appropriate rescaling lemma.

There are many essentially equivalent definitions of a finite-size scaling length; one must choose a definition which is appropriate for the proof at hand. In Chapter 2, we chose a definition of $L_0^*(p)$ for $p > p_c$ and $d = 2$, see (2.13) and (2.14), such that the effective bond density tended exponentially to 1. In fact, in order to do this properly in $d > 2$, one must show instead that the effective density of dual hypersurfaces tends exponentially to 0. Thus, [BCKS] define a finite-size scaling inverse surface tension as

$$A_0(p) = \min\{L^{d-1} \geq 1 : R_{\frac{1}{3},3L} \geq 1 - \epsilon\} \quad \text{for} \quad p > p_c \tag{5.7}$$

with ϵ chosen to satisfy the appropriate rescaling lemma. In $d = 2$, it is possible to show that $A_0(p)$ is indeed equivalent to a fundamental surface tension [BCKS].

We still must define a finite-size scaling correlation length for $p > p_c$. Given the definition (5.5) of the fundamental correlation length, it seems natural to define $L_0(p)$ in terms of a finite-cluster analogue of a rectangle crossing probability. However, [BCKS] find it convenient to consider instead finite-cluster crossings in an annulus

$$H_{L,M} = \mathbb{Z}^d \cap [-L, L + M]^d \setminus (0, M)^d\,,$$

with inner and outer boundaries $\partial_I H_{L,M}$ and $\partial_E H_{L,M}$. [BCKS] say that an occupied cluster $\mathcal{C}_H$ in $H = H_{L,M}$ is H-finite if $H \setminus \mathcal{C}_H$ contains a path—occupied or not—that connects $\partial_I H$ to $\partial_E H$. Let

$$\begin{aligned} S^{\mathrm{fin}}_{L,M}(p) = P_p(\exists \text{ an occupied } H\text{-finite cluster } \mathcal{C}_H \text{ in } H = H_{L,M} \\ \text{that connects } \partial_I H \text{ to } \partial_E H)\,, \end{aligned}$$

with the convention $S^{\text{fin}}_{0,M}(p) = 1$. They define

$$L_0(p) = L_0(p;\epsilon) = 1 + \max\{L \geq 0 \mid S^{\text{fin}}_{L,L}(p) \geq \epsilon\} \qquad \text{for} \quad p > p_c\,, \tag{5.8}$$

and more generally, for $x \geq 1$,

$$L_0(p;\epsilon, x) = 1 + \max\{L \geq 0 \mid S^{\text{fin}}_{L,\lfloor xL \rfloor}(p) \geq \epsilon\} \qquad \text{for} \quad p > p_c\,. \tag{5.9}$$

Note that $L_0(p;\epsilon,x)$ may be finite or infinite, depending on whether or not there exists an $N < \infty$ such that $S^{\text{fin}}_{L,\lfloor xL \rfloor}(p) < \epsilon$ for all $L \geq N$. [BCKS] expect that this definition coincides, say in the sense of Theorem 2.11 (with an $x-$dependent constant c_2, and $c_1(d) = 0$), with the standard correlation length $\xi(p)$ above threshold. However, they are not able to prove this in $d \geq 3$, since the rescaling techniques of [ACCFR] do not work for finite-cluster crossings. In $d = 2$, they prove this using both the RSW and ACCFR lemmas.

While the behavior of $L_0(p)$ below p_c is well understood in general dimension, much less is know about $L_0(p)$ or $A_0(p)$ above p_c. In particular, below p_c, it is straightforward to see that $L_0(p)$ is monotone increasing, left continuous and piecewise constant. Using Theorems 2.9 and 2.11, it is also straightforward to show that

$$L_0(p) \nearrow \infty \quad \text{as} \quad p \nearrow p_c\,,$$

see e.g. [CC1], Proposition 2.11. Furthermore, the jumps in $L_0(p)$ are uniformly bounded (on a logarithmic scale). Namely, by the rescaling inequality

$$R_{\frac{1}{3},6L} \leq \frac{1}{a(d)} R^2_{\frac{1}{3},3L}$$

(which is proved analogously to Lemma 2.3), we have

$$\lim_{\delta \to 0} \frac{L_0(p+\delta)}{L_0(p)} \leq 2\,, \tag{5.10}$$

provided $p < p_c$ and $\epsilon < a(d)$. By contrast, none of these properties are known for $L_0(p)$ above p_c. Therefore [BCKS] turn attention to $A_0(p)$, which they can show is montone decreasing and right continuous. However, in general dimension, they do not have a proof that $A_0(p)$ diverges as p decreases to p_c, nor do they have a bound of the form (5.10). They therefore require several axioms on the behavior of $L_0(p)$ and $A_0(p)$ above p_c.

5.3 The Scaling Axioms and the Results

As mentioned at the beginning of this chapter, the [BCKS] results are established under a set of axioms which they can explictly verify in two dimensions and which they expect to be true whenever the dimension does not exceed the critical dimension d_c (presumably $d_c = 6$). [BCKS] call their axioms the Scaling Axioms since they are to a large extent characterizations of the scaling behaviors implicitly assumed in the physics literature. In this section, we will review the axioms and some of the results from a preliminary version of [BCKS]. Again, this section draws almost verbatim from some of the corresponding sections of [BCKS]. It should be noted that the axioms in the final version of [BCKS] are somewhat different from those presented here: A couple of the axioms here have been shown to be provable from the other axioms and have therefore been removed, and one axiom has been added to obtain stronger results below the scaling window. See [BCKS] for more details.

The Scaling Axioms.

Several of the axioms concern the length scales $L_0(p)$ and $A_0(p)$, and therefore implicitly involve the constant ϵ (see (5.6)-(5.9)). [BCKS] assume that the axioms are true for all $\epsilon < \epsilon_0$, where $\epsilon_0 = \epsilon_0(d)$ depends on the relevant rescaling lemma.

The axioms are written in terms of the equivalence symbol $\asymp$. Here

$$F(p) \asymp G(p)$$

means that there are lower and upper bounds of the form

$$C_1 F(p) \leq G(p) \leq C_2 F(p)$$

where $C_1 > 0$ and $C_2 < \infty$ are constants which do not depend on n or p, as long as p is uniformly bounded away from zero or one, but which may depend on the constants ϵ, $\tilde{\epsilon}$ or x appearing explicitly or implicitly in the axioms. The [BCKS] scaling axioms are

(I) $\quad L_0(p) \to \infty \quad$ as $p \searrow p_c$;

(II) $\quad \displaystyle\limsup_{p \downarrow p_c} \lim_{\delta \to 0} \frac{A_0(p-\delta)}{A_0(p)} < \infty$;

(III) For $0 < \tilde{\epsilon} < \epsilon_0$, $x \geq 1$ and $p > p_c$,

$$A_0(p) \asymp L_0^{d-1}(p) \asymp L_0^{d-1}(p, \tilde{\epsilon}; x);$$

(IV) $\quad \pi_n(p) \asymp \pi_n(p_c) \quad$ if $\quad n \leq L_0(p)$;

(V) $\quad P_{\geq s(n)}(p) \asymp P_{\geq s(n)}(p_c) \quad$ if $\quad n \leq L_0(p)$;

(VI) $\quad \pi_n(p_c) \asymp P_{\geq s(n)}(p_c)$;

(VII) There are finite constants $C_1 > 0$, $C_2 > 0$, $\rho_2 > 0$ and $\rho_1 > \frac{2}{d}$, such that

$$C_1 k^{-1/\rho_1} \leq \frac{\pi_{kn}(p_c)}{\pi_n(p_c)} \leq C_2 k^{-1/\rho_2};$$

(VIII) For $p > p_c$,

$$\chi^{\mathrm{cov}}(p) \asymp \chi^{\mathrm{fin}}(p) \asymp P_\infty^2(p) L_0^d(p);$$

(IX) For $p > p_c$,

$$\pi_{L_0(p)}(p_c) \asymp P_\infty(p).$$

Let us discuss the interpretation of the axioms. The first tells us that the approach to p_c is critical—i.e., continuous or second-order—from above p_c. Recall that the analogue of this for $p \nearrow p_c$ is known rigorously. The second axiom gives a bound of the form (5.10) on the jumps of $A_0(p)$. Note that these axioms are needed due to incomplete knowledge of the continuity properties of the $1/A_0(p)$ and $1/L_0(p)$ above p_c. The third axiom is the assumption of equivalence of length scales above p_c: The second part of it asserts the equivalence of the finite-size scaling lengths at various values of $x \geq 1$ and $\epsilon \in (0, \epsilon_0)$. The first part of it, i.e. $A_0(p) \asymp L_0^{d-1}(p)$, is called Widom scaling. It dimensionally relates the surface tension to the correlation length. Denoting the fundamental surface tension by $\sigma(p)$ and its critical exponent by τ:

$$\sigma(p) \approx |p - p_c|^\tau \quad p \searrow p_c$$

(see [ACCFR] for a defintion of the surface tension), Widom scaling says that

$$\tau = (d-1)\nu,$$

a hyperscaling relation which is expected to hold up to the critical dimension $d = d_c$.

The next two axioms contain a crucial element of the conventional scaling wisdom. Scaling theory asserts that whenever the system is viewed on length scales smaller than the correlation length, it behaves as it does at threshold. Axioms (IV) and (V) assert that this is the case for the connectivity function $\pi(p)$ and the cluster size distribution $P_{\geq s(n)}(p)$.

The sixth axiom is equivalent to the hyperscaling relation

$$d\rho = \delta + 1,$$

as can be seen by using the "definitions" (5.4) and (1.19) of ρ and δ. This of course assumes that the exponents ρ and δ exist. Note that since this is a hyperscaling relation, it is also expected to hold only up to $d = d_c$.

Axiom (VII) implies that the connectivity function $\pi_n(p)$ has upper and lower bounds of power law behavior at threshold. Of course, scaling theory assumes a pure power law with exponent $-1/\rho$. If we assume power laws for χ and L_0 and use the scaling relation implicit in Axioms (VIII) and (IX), the bound $\rho_1 > 2/d$ in Axiom (VII) is equivalent to the very weak bound $\gamma > 0$. (Compare this to the mean field bound $\gamma \geq 1$ in Chapter 3.)

Finally, Axioms (VIII) and (IX) imply scaling and hyperscaling relations among critical exponents, again assuming these exponents exist. Axiom (IX) says that

$$\frac{\nu}{\rho} = \beta,$$

while Axiom (VIII) gives the hyperscaling relations

$$d\nu = 2\beta + \gamma\,.$$

THEOREM 5.1. ([BCKS]) *The Scaling Axioms (I)–(IX) hold in $d = 2$.*

The Results.

In order to state the [BCKS] results, we need to find a scaling window in which the system behaves critically, i.e. an analogue of the function $p(N)$ in the random graph problem. [BCKS] constructed a monotone function $f : R \to [-p_c, 1-p_c]$ with $f(0) = 0$, $f(x) > 0$ for $x > 0$ and $f(x) < 0$ for $x < 0$ such that for

$$p_n = p_c + f\left(\frac{\lambda_n}{n}\right)$$

the system in the finite box $\Lambda_n = B_n(0)$ behaves critically if $\limsup_{n\to\infty} |\lambda_n| < \infty$, subcritically if $\lim_{n\to\infty} \lambda_n = -\infty$ and supercritically if $\lim_{n\to\infty} \lambda_n = +\infty$. Roughly speaking, they find that f is the inverse of the function $x \mapsto 1/L_0(p_c + x)$ and hence $f(x) \approx x^{1/\nu}$. The following theorem makes this notion precise, with part (i) of the theorem giving the scaling behavior of f, and parts (ii)–(iv) giving the expected behavior of the cluster size distribution inside, below and above the window. The symbol $\asymp$ is used in the statement of the theorem, this time for two sequences a_n and b_n of real numbers . We write

$$a_n \asymp b_n$$

if

$$0 < \liminf_{n\to\infty} \frac{a_n}{b_n} \leq \limsup_{n\to\infty} \frac{a_n}{b_n} < \infty\,.$$

The first [BCKS] theorem characterizes the scaling window in terms of the expectation of the largest cluster sizes.

THEOREM 5.2. ([BCKS]) *Suppose that Axioms (I)–(VIII) hold. Let λ_n be a sequence of real numbers, and let $p_n = p_c + f(\lambda_n/n)$.*
(i) $L_0(p_n) \asymp n$ if $\lambda_n \to \lambda_\infty \in \mathbb{R} \setminus \{0\}$, while $L_0(p_n)/n \to 0$ if $\lambda_n \to \pm\infty$.
(ii) If $\limsup_{n\to\infty} |\lambda_n| < \infty$ and $i \in \mathbb{N}$, then

$$E_{p_n}(W^{(i)}_{\Lambda_n}) \asymp s(n)\,.$$

(iii) If $\lim_{n\to\infty} \lambda_n = -\infty$, then

$$\frac{E_{p_n}(W^{(1)}_{\Lambda_n})}{s(n)} \to 0 \qquad as \qquad n \to \infty\,. \tag{5.11}$$

(iv) If $\lim_{n\to\infty} \lambda_n = +\infty$, then

$$\frac{E_{p_n}(W^{(1)}_{\Lambda_n})}{|\Lambda_n| P_\infty(p_n)} \to 1 \qquad as \qquad n \to \infty\,, \tag{5.12}$$

and

$$\frac{E_{p_n}(W^{(2)}_{\Lambda_n})}{E_{p_n}(W^{(1)}_{\Lambda_n})} \to 0 \qquad as \qquad n \to \infty . \tag{5.13}$$

REMARKS. (i) Assume that the critical exponent ν exists (see (1.16)) and that an equivalence of the form of Theorem 2.11 holds for $p > p_c$ as well. Choose $p_n = p_c + f(1/n)$, i.e. $\lambda_n = 1$. Then

$$L_0(p_n) \approx \xi(p_n) \approx |p_n - p_c|^{-\nu} = |f(1/n)|^{-\nu}.$$

By statement (i) of the theorem, $L_0(p_n) \asymp n$, which therefore implies that

$$f(x) \approx \operatorname{sgn}(x)|x|^{1/\nu}.$$

Thus, under the assumption of the existence of ν, the scaling window has width $n^{-1/\nu}$.

(ii) Assuming the existence of the exponent ρ, see (5.4), the theorem implies that inside the scaling window the largest, second largest, third largest, $\cdots$ clusters scale like n^{d_f}, with $d_f = d - 1/\rho$. On the other hand, below the scaling window the size of the largest cluster (and hence of all clusters) goes to zero on the scale n^{d_f}, while above the scaling window the largest cluster becomes dominant and its size tends to 1 on the scale $n^d P_\infty$. Above the window, it is also possible to show that the size of the largest cluster tends to infinity on the scale n^{d_f}, although this requires Axiom (IX) (which was not used in Theorem 5.2), in addition to the theorem. Together with Axiom (VII) and statement (i) of the theorem, Axiom (IX) implies that

$$\frac{|\Lambda_n| P_\infty(p_n)}{s(n)} = \frac{P_\infty(p_n)}{\pi_n(p_c)} \asymp \frac{\pi_{L_o(p_n)}(p_c)}{\pi_n(p_c)} \to \infty .$$

So, by (5.12), the size of the largest cluster diverges on the scale $s(n) \approx n^{d_f}$.

[BCKS] also prove analogues of statements (ii)–(iv) of the theorem for convergence in probability, rather than in expectation. Morever, they show that the distribution of $W^{(i)}_{\Lambda_n}$ is nontrivial within the scaling window. These results require some delicate second moment estimates which are beyond the scope of theses notes. Also, under an additional axiom, [BCKS] establish that $W^{(1)}_{\Lambda_n}$ scales like $s(L_o(p_n)) \log[n/L_o(p_n)]$ below the scaling window, again both in expectation and in probability. This also requires some very delicate estimates. The reader is referred to [BCKS] for precise statements of these results and for their proofs.

One final result is worth mentioning, since it is used in the proofs of the other results and is of interest in its own right. It concerns the number of clusters on scales $m < n$. Before stating the result, it should be noted that, due to (5.11), the "incipient infinite cluster" inside the scaling window is not unique, in the sense that $W^{(2)}_{\Lambda_n}$ is of the same scale as $W^{(1)}_{\Lambda_n}$. This should be contrasted with the behavior (5.13) of $W^{(2)}_{\Lambda_n}/W^{(1)}_{\Lambda_n}$ above the scaling window, a remnant of the uniqueness of the infinite cluster above p_c. The next theorem relates the non-uniqueness of the "incipient infinite cluster" inside the scaling window to the property of scale invariance at p_c. Basically, it says that the number of clusters of scale m in a system of scale n is a function only of the ratio n/m. How can this hold on *all* scales m? The only way it can be true is if the system has a fractal-like structure with smaller clusters inside holes in larger clusters. The theorem concerns the numbers $N_\Lambda(s_1, s_2)$ and $\tilde{N}_\Lambda(s_1, s_2)$ of clusters with size between s_1 and s_2, see (5.2) and (5.3).

THEOREM 5.3. ([BCKS]) *Suppose that Axioms (IV)–(VII) hold. Then there are strictly positive, finite constants C_1, C_2 and k_0 such that*

$$C_1\left(\frac{n}{m}\right)^d \leq E_p\big(\tilde{N}_\Lambda(s(m), s(km))\big) \leq E_p\big(N_\Lambda(s(m), s(km))\big) \leq C_2\left(\frac{n}{m}\right)^d$$

provided $\Lambda = \Lambda_n$, $k_0 m \leq \min\{L_0(p), n\}$ and $k \geq k_0$.

PROOF OF THEOREM 5.3. We follow essentially verbatim the proof of [BCKS]. We begin with the fact that for an arbitrary configuration ω, the number of clusters of size s can be rewritten as

$$|\{i | W_\Lambda^{(i)} = s\}| = \sum_{x\in\Lambda} \frac{1}{s} \mathbf{1}_{\{|\mathcal{C}_\Lambda(x)|=s\}},$$

where, as usual, $\mathbf{1}_A$ denotes the indicator of the event A. As an immediate consequence,

$$E_p(N_\Lambda(s_1, s_2)) = \sum_{s=s_1}^{s_2} \sum_{x\in\Lambda} \frac{1}{s} P_p(|\mathcal{C}_\Lambda(x)| = s). \tag{5.14}$$

In a similar way,

$$E_p(\tilde{N}_\Lambda(s_1, s_2)) = \sum_{s=s_1}^{s_2} \sum_{x\in\Lambda} \frac{1}{s} P_p(|\mathcal{C}_\Lambda(x)| = s,\ x \not\leftrightarrow \partial\Lambda). \tag{5.15}$$

First we prove the upper bound, which follows easily from the representation above and the axioms. Choosing $s_1 = s(m)$ and $s_2 = s(km)$, using the representation (5.14) and bounding the factor $1/s$ in (5.14) by $1/s_1 = 1/s(m)$, we get

$$\begin{aligned} E_p\big(N_\Lambda(s(m), s(km))\big) &\leq \frac{1}{s(m)} \sum_{x\in\Lambda} \sum_{s\geq s(m)} P_p(|\mathcal{C}_\Lambda(x)| = s) \\ &= \frac{1}{s(m)} \sum_{x\in\Lambda} P_p(|\mathcal{C}_\Lambda(x)| \geq s(m)) \\ &\leq \frac{(2n)^d}{s(m)} P_{\geq s(m)}(p), \end{aligned} \tag{5.16}$$

where in the last step we used the definition (1.8) of $P_{\geq s(m)}(p)$ and the fact that $|\mathcal{C}_\Lambda(x)| \leq |\mathcal{C}(x)|$. Since $k_0 m \leq L_0(p)$ implies $m \leq L_0(p)$, we may use Axioms (V) and (VI) to bound the right hand side of (5.16). We get

$$\frac{(2n)^d}{s(m)} P_{\geq s(m)}(p) \leq C_2 \frac{(2n)^d}{s(m)} \pi_m(p_c) = C_2 \left(\frac{n}{m}\right)^d \tag{5.17}$$

where C_2 is a finite constant. (5.16) and (5.17) imply the upper bound.

The lower bound is somewhat more difficult. To prove it, we choose a constant $\tilde{k}_0$ to be fixed in the course of the proof. Given $\tilde{k}_0$, we choose $k_0 \geq \tilde{k}_0$ in such a way that the assumption $k \geq k_0$ implies $s(km) \geq s(\tilde{k}_0 m)$ (the existence of such a k_0 is guaranteed by Axiom VII). Using this bound and (5.15), we have

$$\begin{aligned} E_p\big(\tilde{N}_\Lambda(s(m), s(km))\big) &\geq E_p\big(\tilde{N}_\Lambda(s(m), s(\tilde{k}_0 m) - 1)\big) \\ &\geq \sum_{s=s(m)}^{s(\tilde{k}_0 m)-1} \sum_{x \in \Lambda_{\frac{n}{2}}} \frac{1}{s} P_p(|\mathcal{C}_\Lambda(x)| = s, x \not\leftrightarrow \partial\Lambda) \\ &= \sum_{s=s(m)}^{s(\tilde{k}_0 m)-1} \sum_{x \in \Lambda_{\frac{n}{2}}} \frac{1}{s} P_p(|\mathcal{C}(x)| = s, x \not\leftrightarrow \partial\Lambda), \end{aligned} \tag{5.18}$$

where in the second step we bounded the sum over $\Lambda = \Lambda_n$ from below by a sum over $\Lambda_{\frac{n}{2}}$. The fact that x is not connected to $\partial\Lambda$ allows us to replace $\mathcal{C}_\Lambda(x)$ by $\mathcal{C}(x)$ in the last step. Bounding the factor $1/s$ in (5.18) from below by $1/s(\tilde{k}_0 m)$, we get

$$\begin{aligned} &E_p\big(\tilde{N}_\Lambda(s(m), s(km))\big) \\ &\qquad \geq \frac{1}{s(\tilde{k}_0 m)} \sum_{x \in \Lambda_{\frac{n}{2}}} P_p(s(m) \leq |\mathcal{C}(x)| < s(\tilde{k}_0 m), x \not\leftrightarrow \partial\Lambda) \\ &\qquad = \frac{1}{s(\tilde{k}_0 m)} \sum_{x \in \Lambda_{\frac{n}{2}}} \Big[P_p(s(m) \leq |\mathcal{C}(x)| < s(\tilde{k}_0 m)) \\ &\qquad\qquad - P_p(s(m) \leq |\mathcal{C}(x)| < s(\tilde{k}_0 m), x \leftrightarrow \partial\Lambda)\Big] \\ &\qquad \geq \frac{1}{s(\tilde{k}_0 m)} \sum_{x \in \Lambda_{\frac{n}{2}}} \Big[P_p(s(m) \leq |\mathcal{C}(x)| < s(\tilde{k}_0 m)) - \pi_{n/2}(p)\Big] \\ &\qquad = \frac{n^d}{s(\tilde{k}_0 m)} \Big[P_{\geq s(m)}(p) - P_{\geq s(\tilde{k}_0 m)}(p) - \pi_{n/2}(p)\Big]. \end{aligned}$$

In the third step above, we relaxed the condition that $s(m) \leq |\mathcal{C}(x)| < s(\tilde{k}_0 m)$ in the second term in the sum. Since $n \geq k_0 m \geq \tilde{k}_0 m$ by the assumption $k_0 m \leq \min\{L_0(p), n\}$, we obtain

$$E_p\big(\tilde{N}_\Lambda(s(m), s(km))\big) \geq \frac{n^d}{s(\tilde{k}_0 m)} \Big[P_{\geq s(m)}(p) - P_{\geq s(\tilde{k}_0 m)}(p) - \pi_{\tilde{k}_0 m/2}(p)\Big].$$

Again by the assumption $mk_0 \le \min\{L_0(p), n\}$, we have $\tilde{k}_0 m \le L_0(p)$. We therefore may use Axioms (IV), (V) and (VI) in conjunction with the bound $\pi_{\tilde{k}_0 m}(p) \le \pi_{\tilde{k}_0 m/2}(p)$ to conclude that

$$E_p\big(\tilde{N}_\Lambda(s(m), s(km))\big) \ge \frac{n^d}{s(\tilde{k}_0 m)} \Big[\tilde{C}_1 \pi_m(p_c) - \tilde{C}_2 \pi_{\tilde{k}_0 m/2}(p_c)\Big] \,,$$

where $\tilde{C}_1$ and $\tilde{C}_2$ are strictly positive, finite constants. Finally using Axiom (VII) to bound $\pi_{\tilde{k}_0 m/2}(p_c)$ from above by a small constant times $\pi_m(p_c)$, we get

$$E_p\big(\tilde{N}_\Lambda(s(m), s(km))\big) \ge \frac{n^d}{s(\tilde{k}_0 m)} \tfrac{1}{2}\tilde{C}_1 \pi_m(p_c) \,,$$

provided $\tilde{k}_0$ is chosen large enough (depending on the constants in Axiom (VII) and the ratio of $\tilde{C}_1$ and $\tilde{C}_2$). We finally use that $s(\tilde{k}_0 m) \le \tilde{k}_0^d s(m)$ to conclude that

$$E_p\big(\tilde{N}_\Lambda(s(m), s(km))\big) \ge C_1 \frac{n^d}{s(m)} \pi_m(p_c) = C_1 \left(\frac{n}{m}\right)^d \,,$$

where $C_1 = \frac{1}{2}\tilde{k}_0^d \tilde{C}_1$. This proves the lower bound. $\square$

5.4 Interpretation of the Results

Assuming the existence of critical exponents and monotonicity of various quantities, the results say basically that the scaling window is of the form

$$p_n = p_c \pm \frac{c}{n^{1/\nu}},$$

that inside the window

$$W^{(1)} \approx n^{d_f},$$
$$W^{(2)} \approx n^{d_f},$$
$$\vdots$$

while above the window

$$W^{(1)} \approx n^d P_\infty \,,$$
$$W^{(1)}/n^{d_f} \to \infty \,,$$
$$W^{(2)}/W^{(1)} \to 0,$$

and below the window

$$W^{(1)}/n^{d_f} \to 0$$

where, in fact,

$$W^{(1)} \approx \xi^{d_f} \log n/\xi \,.$$

These results hold both in expectation and in probability.

How can we understand the form of the window? As explained earlier, the system is expected to behave critically whenever the length scale is less than the correlation length. Indeed, this is the content of Axioms (IV) and (V). But this means that

$$n \leq \tilde{\lambda}\xi \approx \tilde{\lambda}|p-p_c|^{-\nu},$$

i.e.

$$p \approx p_c \pm \frac{\lambda}{n^{1/\nu}},$$

where $\tilde{\lambda}$ and λ are constants. This is of course precisely the content of Theorem 5.2(i) (and Remark (i) following the theorem).

What would these results say if we attempted to apply them in the case of random graph model (to which they of course do not rigorously apply)? Let us use the hyperscaling relation

$$d\nu = \gamma + 2\beta$$

and the observation that the volume N of our system is just n^d, to rewrite the window in the form

$$\begin{aligned} p_n &= p_c \pm \frac{\lambda}{n^{1/\nu}} \\ &= p_c\left(1 \pm \frac{c}{n^{1/\nu}}\right) \\ &= p_c\left(1 \pm \frac{c}{N^{1/d\nu}}\right) \\ &= p_c\left(1 \pm \frac{c}{N^{1/(\gamma+2\beta)}}\right). \end{aligned} \tag{5.19}$$

Similarly, let us use the hyperscaling relation

$$d_f/d = \delta/(1+\delta)$$

to rewrite the size of the largest cluster as

$$\begin{aligned} W^{(1)} &\approx n^{d_f} \\ &\approx N^{d_f/d} \\ &\approx N^{\delta/(1+\delta)}. \end{aligned} \tag{5.20}$$

Noting that the random graph model is a mean-field model, we expect (and in fact it can be verified) that $\gamma = 1$, $\beta = 1$ and $\delta = 2$. Using also $p_c = 1/N$, we have a window of the form

$$p(N) = \frac{1}{N} \pm \frac{c}{N^{4/3}},$$

and within that window

$$W^{(1)} \approx N^{2/3},$$

just the values obtained in the combinatoric calculations on the random graph model.

The results also have implications for finite-size scaling. Indeed, the form (5.19) of the window tells us precisely how to locate the critical point, i.e. it tells us the correct region about p_c in which to do critical calculations. Similarly, (5.20) tells us how to extrapolate the scaling of clusters in the critical regime.

Finally, the results tell us that we may use the largest cluster in the box as a candidate for the incipient infinite cluster. Within the window, it is not unique, in the sense that there are many clusters of this scale. However, outside the window (even including a region where p is not strictly greater than p_c as $n \to \infty$), there is a unique cluster of largest scale. This is the analogue of what is called the *dominant component* in the random graph problem.

CHAPTER VI

THE BK(R) INEQUALITY

In this chapter, we review D. Reimer's beautiful new proof of the BK inequality for arbitrary events in percolation. The proof given here is based on a copy of a preliminary manuscript and some notes by D. Reimer [Re], as well as on a lecture by J. Kahn and on some comments on a preliminary version of this chapter by C. Borgs, H. Kesten and P. Deligne. We have modified some of Reimer's notation and added a few details to the proofs we have seen, but the main proof presented here is very similar to that given by Reimer. In particular, the notion of butterflies is due to him, and it is this notion that he uses in his proof of the main lemma (Lemma 6.2). It should be noted that a form of the proof of the sufficiency of the main lemma was already known to van den Berg [Be] and Talagrand [Ta].

We begin the chapter with a section on various equivalent forms of the BK inequality, one of which—the Fishburn-Shepp [FiS] form—is the one ultimately proved by Reimer. Throughout the chapter, we restrict attention to independent percolation on a bond (or site) lattice $\mathbb{B}$ with $|\mathbb{B}| = n < \infty$. Here $\Omega = \{0,1\}^{\mathbb{B}}$, so that $|\Omega| = 2^n$.

6.1 Equivalent Forms of the Inequality

First we reformulate the disjoint occurence event $A \circ B$ in terms of cylinders. Given a configuration $\omega \in \Omega$ and a set of bonds $S \subset \mathbb{B}$, we define the *cylinder* $[\omega]_S$ by

$$[\omega]_S = \{\omega' : \omega'_b = \omega_b \quad \forall b \in S\}.$$

With this definition, we may rewrite $A \circ B$ as

$$A \circ B = \{\omega : \exists S_A, S_B \subset \mathbb{B},\ S_A \cap S_B = \emptyset,\ [\omega]_{S_A} \subset A,\ [\omega]_{S_B} \subset B\}. \tag{6.1}$$

As we saw in Chapter 1, the original form of the BK inequality is

$$P_p(A \circ B) \leq P_p(A)\, P_p(B) \tag{6.2}$$

for any events A and B and for all $0 \leq p \leq 1$. The first simplification was due to van den Berg and Feibig [BF] who showed that it is sufficient to prove the inequality for $p = 1/2$. We will not give the proof here, although we will note that it is not terribly difficult. It relies on a general lemma that, roughly speaking, relates measures that respect the same notion of disjointness. This simplification is significant in that the inequality now becomes a purely combinatorial one. Since each bond is equally likely to open or closed, the probability of any event A occuring is just $|A|/|\Omega|$. In other words, the BF form of the BK inequality is given by

$$|A \circ B|\,|\Omega| \le |A|\,|B| \tag{6.3}$$

for any events A and B.

Fishburn and Shepp [FiS] discovered another way of expressing the BK inequality. It is, in fact, this form of the inequality that Reimer [Re] proved. The FS form of the BK inequality is

$$m\,|\Omega| \le \left|\bigcup_{i=1}^{m}[\omega^i]_{S_i}\right|\,\left|\bigcup_{i=1}^{m}[\omega^i]_{S_i^c}\right| \tag{6.4}$$

for all $1 \le m \le |\Omega|$, all distinct configurations $\omega^1, \dots, \omega^m \in \Omega$, and all sets $S_1, \dots, S_m \subset \mathbb{B}$. We will now show that (6.4) is equivalent to (6.2).

Proof that (6.2) implies (6.4). Let $1 \le m \le |\Omega|$, let $\omega^1, \dots, \omega^m \in \Omega$ be distinct configurations, and let $S_1, \dots, S_m \subset \mathbb{B}$. Consider

$$A = \bigcup_{i=1}^{m}[\omega^i]_{S_i} \quad \text{and} \quad B = \bigcup_{i=1}^{m}[\omega^i]_{S_i^c}.$$

Notice that

$$\{\omega^1, \dots, \omega^m\} \subset A \circ B\,.$$

Hence, by (6.2),

$$\begin{aligned} m\,|\Omega| &\le |A \circ B|\,|\Omega| \\ &\le |A|\,|B| \\ &= \left|\bigcup_{i=1}^{m}[\omega^i]_{S_i}\right|\,\left|\bigcup_{i=1}^{m}[\omega^i]_{S_i^c}\right|. \end{aligned}$$

□

Proof that (6.4) implies (6.2). We will only show that (6.4) implies (6.3), and rely on the equivalence of (6.2) and (6.3). Let $A, B \subset \Omega$. Let $m = |A \circ B|$. Let $\omega^1, \dots, \omega^m$ denote the distinct elements of $A \circ B$. By the definition (6.2) of $A \circ B$, there are S_i such that

$$[\omega^i]_{S_i} \subset A \text{ and } [\omega^i]_{S_i^c} \subset B.$$

So, by (6.4),

$$\begin{aligned} |A \circ B|\,|\Omega| &= m\,|\Omega| \\ &\le \left|\bigcup_{i=1}^{m}[\omega^i]_{S_i}\right|\,\left|\bigcup_{i=1}^{m}[\omega^i]_{S_i^c}\right| \\ &\le |A|\,|B|. \end{aligned}$$

□

6.2 Preliminaries to the Proof of the BK Inequality

We begin the proof with some notation and a couple of simple observations. In the rest of this chapter, we will depart from our previous convention and use w, x, y and z to denote configurations, and i and j to denote sites or bonds. The symbol b will be reserved for butterflies, to be defined below.

Let $x, y \in \Omega$. The *antipode* x' of the configuration x is given by $x'_i = 1 - x_i$, and the antipode Q' of the event $Q \subset \Omega$ is $Q' = \{x : x' \in Q\}$. The *subcube* generated by two configurations $x, y \in \Omega$ is $\langle x, y\rangle = [x]_{\{i : x_i = y_i\}} = [y]_{\{i : x_i = y_i\}}$. Notice that subcubes are equivalent to cylinders.

Reimer's fundamental construct in the proof of the BK inequality is what he called a *butterfly*:

DEFINITION: A *butterfly* b on Ω is an ordered pair $b = (x, y)$ with $x, y \in \Omega$. A butterfly generates the following four subsets of Ω:

$$\begin{aligned} \text{body}(x,y) &= \{x\} \\ \text{tip}(x,y) &= \{y\} \\ \text{right wing}(x,y) &= r(b) = \langle x, y\rangle \\ \text{left wing}(x,y) &= \ell(b) = \langle x, y'\rangle\,. \end{aligned}$$

Notice that

$$A \circ B = \{x : \exists \text{ butterfly } b,\ \ \text{body}(b) = x,\ r(b) \subset A,\ \ell(b) \subset B\}.$$

DEFINITION: A *flock* B of butterflies on Ω is a set $B = \{b : b \in B\}$ of butterflies on Ω. It also generates four subsets of Ω:

$$\begin{aligned} \text{body}(B) = X(B) &= \bigcup_{b \in B} \text{body}(b) \\ \text{tip}(B) = Y(B) &= \bigcup_{b \in B} \text{tip}(b) \\ \text{right wing}(B) = R(B) &= \bigcup_{b \in B} r(b) \\ \text{left wing}(B) = L(B) &= \bigcup_{b \in B} \ell(b)\,. \end{aligned}$$

Notice that the BK inequality in the Fishburn-Shepp form (6.4) is equivalent to the statement that

$$|R(B)|\,|L(B)| \geq |\Omega|\,|X(B)| \tag{6.5}$$

for all flocks B of butterflies with distinct bodies. (To see the equivalence, consider the correspondence that takes a configuration $x \in \Omega$ and a set of bonds $S \subset \mathbb{B}$ and generates a butterfly $(x, \hat{x})$ with $x_i = \hat{x}_i$ iff $i \in S$.)

Now consider a cylinder $A = [y]_S$. Given a configuration x, we define the *antipode of x in A* by

$$x_i'^{(A)} = \begin{cases} x_i & \text{if } i \in S \\ 1 - x_i & \text{if } i \in S^c \,. \end{cases}$$

Thus $x'^{(A)}$ is the configuration which agrees with x on those sites i on which A is fixed and differs from x on the sites i on which A varies. Notice that

$$A = \langle x, y \rangle \iff y = x'^{(A)}. \tag{6.6}$$

Given the notion of an antipode, we now define, for each cylinder A and each butterfly (x, y) with $x \in A$ and $y \in A$, the left and right wing of (x, y), considered as a butterfly in A. Namely, we define

$$\begin{aligned} r_A(x, y) &= \langle x, y \rangle \quad \text{and} \\ \ell_A(x, y) &= \langle x, y'^{(A)} \rangle. \end{aligned}$$

The next lemma defines what we will call subbutterflies on cylinders.

LEMMA 6.1. (SUBBUTTERFLY LEMMA) *Let $A \subset \Omega$ be a cylinder and let (x, y) be a butterfly on Ω with $x \in A$. Then there exists a unique $\tilde{y} \in A$ such that*

$$\begin{aligned} r_A(x, \tilde{y}) &= r(x, y) \cap A \quad \text{and} \\ \ell_A(x, \tilde{y}) &= \ell(x, y) \cap A. \end{aligned}$$

PROOF. Let the set S be defined by $A = [x]_S$. If $\tilde{y}$ is defined by $\langle x, \tilde{y} \rangle = \langle x, y \rangle \cap A$, then it is easy to see that

$$\tilde{y}_i = \begin{cases} x_i & \text{if } i \in S \\ y_i & \text{if } i \in S^c. \end{cases}$$

Now, it simply remains to check that $\langle x, \tilde{y}'^{(A)} \rangle = \langle x, y' \rangle \cap A$ for this choice of $\tilde{y}$. $\square$

This allow us also to define subflocks on cylinders:

DEFINITION: Let $b = (x, y)$ with x in some cylinder A. We denote by b_A the *subbutterfly* $(x, \tilde{y})$ on A, with $\tilde{y} \in A$ as given in the Subbutterfly Lemma. Let B be a flock of butterflies on Ω. We define the *subflock* $B_A = B_A(B)$ as the flock on A given by

$$B_A = \{ b_A : b \in B \;\; \text{body}(b) \in A \}.$$

We also define $X_A(B)$, $Y_A(B)$, $R_A(B)$, and $L_A(B)$ in the obvious way; we suppress B when it is convenient.

Notice that $X_A = X \cap A$ and that

$$\begin{aligned} R_A &= \bigcup_{b_A \in B_A} r_A(b_A) \;=\; R \cap A \;\subset R \\ L_A &= \bigcup_{b_A \in B_A} \ell_A(b_A) \;=\; L \cap A \;\subset L\,. \end{aligned}$$

6.3 The Proof of the BK Inequality

We now come to the main lemma in the proof of the BK inequality.

LEMMA 6.2. *For any flock of butterflies B,*

$$|R \cap L'| \geq |X|.$$

Before proving the lemma, let us see how to apply it to prove the BK inequality.

PROOF OF THEOREM 1.4 GIVEN LEMMA 6.2. It suffices to show the Fishburn-Shepp form (6.5) of the BK inequality. Let B be a flock of butterflies with distinct bodies. Partitioning according to subcubes and using the fact that $R_A \subset R$ and $L_A \subset L$ for a subflock on any subcube A, we have

$$\begin{aligned} |R|\,|L| &= |\{(z,w) \in R \times L\}| \\ &= \sum_A |\{(z,w) \in R \times L : \langle z,w \rangle = A\}| \\ &\geq \sum_A |\{(z,w) \in R_A \times L_A : \langle z,w \rangle = A\}|. \end{aligned}$$

But by (6.6), if $\langle z,w \rangle = A$, then $z = w'^{(A)}$, so that $w \in L_A$ implies $z \in L_A'^{(A)}$. Given that $z \in R_A$ also, it follows that $z \in R_A \cap L_A'^{(A)}$. Since, for a given A and z, the condition $\langle z,w \rangle = A$ uniquely specifies w, we have

$$|\{(z,w) \in R_A \times L_A : \langle z,w \rangle = A\}| = |R_A \cap L_A'^{(A)}|.$$

Thus applying the main lemma on the state space A, we have

$$|R|\,|L| \geq \sum_A |X_A|. \tag{6.7}$$

Now an easy counting argument gives that the right hand side of (6.7) is equal to $|X||\Omega|$. Indeed,

$$\sum_A |X \cap A| = \sum_A \sum_{x \in X} |\{x\} \cap A| = \sum_{x \in X} \sum_{A \ni x} 1 = \sum_{x \in X} |\Omega| = |X||\Omega|,$$

which, together with (6.7), is the desired result. $\square$

Now we come to the heart of the matter: Reimer's proof of the main lemma.

PROOF OF LEMMA 6.2. First, we note that the statement of the lemma is obviously equivalent to the statement that $|R \cap L| \geq |Y|$ for all flocks of butterflies B. In order to see this, just interchange all tips and bodies in the butterflies of the flock B. Since this leaves the right wings invariant, but changes the set L to L' and the set X to Y, the claim follows. Next we note that it is enough to prove the statement $|R \cap L| \geq |Y|$ for all flocks of butterflies with distinct tips, since this immediately implies the same statement for general flocks of butterflies. But for flocks of butterflies with distinct tips, we have $|Y| = |B|$, so that suffices to prove $|R \cap L| \geq |B|$. Using de Morgan's laws, this is equivalent to

$$|R^c \cup L^c| \leq |\Omega| - |B|,$$

or

$$|R^c| + |R \cap L^c| + |B| \leq |\Omega| = 2^n \tag{6.8}$$

To obtain (6.8), we will construct three injective maps:

$$\begin{aligned} \alpha &: R^c \to \mathbb{R}^{2^n} \\ \beta &: R \cap L^c \to \mathbb{R}^{2^n} \\ \gamma &: B \to \mathbb{R}^{2^n}. \end{aligned}$$

We will show that the images of these maps are disjoint and that the union of the images is a set of linearly independent vectors in $\mathbb{R}^{2^n}$. This immediately implies that the number of elements in the union, and hence on the left hand side of (6.8), is bounded above by 2^n. It suffices to verify the following six statements:

(1) $\alpha(R^c) \perp \beta(R \cap L^c)$
(2) $\alpha(R^c) \perp \gamma(B)$
(3) $\beta(R \cap L^c) \perp \gamma(B)$
(4) $\alpha(R^c)$ is linearly independent
(5) $\beta(R \cap L^c)$ is linearly independent
(6) $\gamma(B)$ is linearly independent.

We begin by defining the maps. To define α on R^c, first define α on $\{0,1\}$:

$$\alpha(x_i) = \begin{cases} (1,1) & \text{if } x_i = 0 \\ (0,1) & \text{if } x_i = 1. \end{cases}$$

Now, to extend α to $\Omega = \{0,1\}^{\mathbb{B}}$, we must first set some notation. Let $\oplus$ denote concatenation given by, $(a,b) \oplus (c,d) = (a,b,c,d)$. Let $\otimes$ be the tensor product given by $(a,b) \otimes v = av \oplus bv$ for $a, b \in \mathbb{R}$ and $v \in \mathbb{R}^m$. Equipping $\mathbb{R}^{2^n}$ with the standard inner product: $\langle v \,|\, w \rangle = \sum_{i=1}^{2^n} v_i w_i$, notice that a easy inductive proof yields

$$\langle \bigotimes_{i=1}^n v_i \,|\, \bigotimes_{i=1}^n w_i \rangle = \prod_{i=1}^n \langle v_i \,|\, w_i \rangle \tag{6.9}$$

for $v_i, w_i \in \mathbb{R}^2, 1 \leq i \leq n$. With notation in hand, let

$$\alpha(x) = \bigotimes_{i=1}^n \alpha(x_i)$$

for each $x \in \Omega$. Similarly, we define $\beta(x)$ on Ω by

$$\beta(x) = \bigotimes_{i=1}^n \beta(x_i)$$

where

$$\beta(x_i) = \begin{cases} (1,0) & \text{if } x_i = 0 \\ (1,-1) & \text{if } x_i = 1\,, \end{cases}$$

and we define $\gamma(x,y)$ on $\Omega \times \Omega$ by

$$\gamma(x,y) = \bigotimes_{i=1}^{n} \gamma((x_i,y_i))$$

where

$$\gamma(x_i,y_i) = \begin{cases} (1,0) & \text{if } (x_i,y_i) = (0,0) \\ (1,1) & \text{if } (x_i,y_i) = (0,1) \\ (0,1) & \text{if } (x_i,y_i) = (1,0) \\ (1,-1) & \text{if } (x_i,y_i) = (1,1). \end{cases}$$

It remains to verify the six conditions that imply the linear independence. The functions have been cleverly defined so that most of this will be quite routine.

(1) $\alpha(R^c) \perp \beta(R \cap L^c)$

If $x \in R^c$ and $y \in R \cap L^c$, then $x \neq y$ and so $x_i \neq y_i$ for some i. So either

$$\langle \alpha(x_i) \,|\, \beta(y_i) \rangle = \langle (1,1) \,|\, (1,-1) \rangle = 0$$

or

$$\langle \alpha(x_i) \,|\, \beta(y_i) \rangle = \langle (0,1) \,|\, (1,0) \rangle = 0.$$

Hence, recalling (6.9), we have that

$$\langle \alpha(x) \,|\, \beta(y) \rangle = 0.$$

Since it is easy to see that neither $\alpha(x)$ nor $\beta(y)$ can be the zero vector, it follows that $\alpha(R^c) \perp \beta(R \cap L^c)$.

(2) $\alpha(R^c) \perp \gamma(B)$

If $z \in R^c$ and $(x,y) \in B$, then $z \notin \langle x,y \rangle$ and so $x_i = y_i \neq z_i$ for some i. So either

$$\langle \alpha(z_i) \,|\, \gamma(x_i,y_i) \rangle = \langle (0,1) \,|\, (1,0) \rangle = 0$$

or

$$\langle \alpha(z_i) \,|\, \gamma(x_i,y_i) \rangle = \langle (1,1) \,|\, (1,-1) \rangle = 0.$$

Hence $\langle \alpha(z) \,|\, \gamma(x,y) \rangle = 0$, which yields $\alpha(R^c) \perp \gamma(B)$.

(3) $\beta(R \cap L^c) \perp \gamma(B)$

If $z \in R \cap L^c$ and $(x,y) \in B$, then $z \notin \langle x,y' \rangle$ and so $x_i = 1 - y_i \neq z_i$ for some i. So either

$$\langle \beta(z_i) \,|\, \gamma(x_i,y_i) \rangle = \langle (1,0) \,|\, (0,1) \rangle = 0$$

or

$$\langle \beta(z_i) \,|\, \gamma(x_i, y_i) \rangle = \langle (1,-1) \,|\, (1,1) \rangle = 0.$$

Hence $\langle \beta(z) \,|\, \gamma(x,y) \rangle = 0$, which yields $\beta(R \cap L^c) \perp \gamma(B)$.

(4) $\alpha(R^c)$ is linearly independent

We will show that $\alpha(\Omega)$ is linearly independent. Here we could rely on the obvious implication that since $\{(1,1),(0,1)\}$ is a basis of $\mathbb{R}^2$, the tensor product $\alpha(\Omega) = \otimes_{i=1}^{n}\{(1,1),(0,1)\}$ must be a basis of $\mathbb{R}^{2^n}$. Instead, let us show this explicitly, in the process developing a method which will apply also to the less obvious case of independence of $\gamma(B)$. For $0 \le k < 2^n$, let x^k be the configuration in Ω given by the binary representation of k so that $\Omega = \{x^k : 0 \le k < 2^n\}$. If we let $A^{(n)}$ be the $2^n \times 2^n$ matrix formed by letting row $k+1$ be the vector $\alpha(x^k)$, then we wish to show that $\det A^{(n)} \neq 0$. Indeed we will show $\det A^{(n)} = 1$ by induction. The case $n = 1$ is trivial. Assume that $\det A^{(n)} = 1$. Noting that the first 2^n configurations begin with a 0 and the next 2^n begin with a 1, and using the definition of α, it is easy to see that for $0 \le k < 2^n$,

$$\begin{aligned} A^{(n+1)}_{k+1,\cdot} &= (1,1) \otimes \bigotimes_{i=2}^{n+1} \alpha(x^k(i)) \\ &= \bigotimes_{i=2}^{n+1} \alpha(x^k(i)) \oplus \bigotimes_{i=2}^{n+1} \alpha(x^k(i)) \\ &= A^{(n)}_{k+1,\cdot} \oplus A^{(n)}_{k+1,\cdot} \ , \end{aligned}$$

while for $2^n \le k < 2^{n+1}$,

$$\begin{aligned} A^{(n+1)}_{k+1,\cdot} &= (0,1) \otimes \bigotimes_{i=2}^{n+1} \alpha(x^k(i)) \\ &= \bigoplus_{j=1}^{2^n} 0 \oplus \bigotimes_{i=2}^{n+1} \alpha(x^k(i)) \\ &= \bigoplus_{j=1}^{2^n} 0 \oplus A^{(n)}_{k+1,\cdot} \ . \end{aligned}$$

That is,

$$A^{(n+1)} = \begin{pmatrix} A^{(n)} & A^{(n)} \\ 0 & A^{(n)} \end{pmatrix}$$

so that $\det A^{(n+1)} = \det A^{(n)} \det A^{(n)} = 1$, completing our inductive proof. As mentioned above, this method will serve us for the final two arguments, of which the latter is the more complicated one.

(5) $\beta(R \cap L^c)$ is linearly independent

This proof is exactly like the last proof. Here, the recursion from the matrix $A^{(n+1)}$ to $A^{(n)}$ is

$$A^{(n+1)} = \begin{pmatrix} A^{(n)} & 0 \\ A^{(n)} & -A^{(n)} \end{pmatrix}$$

which again gives $\det A^{(n+1)} \neq 0$.

(6) $\gamma(B)$ is linearly independent

For this argument, it is sufficient to prove the independence on $\mathbb{Z}_2^{2^n}$ rather than $\mathbb{R}^{2^n}$, and, as will become clear, it turns out to be much simpler for $\mathbb{Z}_2^{2^n}$. For the moment, simply note that, in $\mathbb{Z}_2^2$, $\gamma((1,1)) = (1,1)$, so that $\gamma((1,1)) = \gamma((0,1))$.

Notice that since the tips are distinct (recall that we made this assumption in the first paragraph of the proof), for each $(x, y) \in B$, there is a function $x : \Omega \to \Omega$ such that $(x, y) = (x(y), y)$. This in turn induces a function $\gamma_x(y) : \Omega \to \mathbb{R}^{2^n}$ (or $\mathbb{Z}_2^{2^n}$) which coincides with γ on $\mathbb{R}^{2^n}$ (or $\mathbb{Z}_2^{2^n}$):

$$\gamma_x(y) = \bigotimes_{i=1}^{n} \gamma(x(y)_i, y_i).$$

Again denoting by y^k the binary representation of k, and now writing rows and columns of matrices as arguments rather than as subscripts, it suffices to show that for all functions $x : \Omega \to \Omega$, the matrix $A_x^{(n)}$ defined by

$$A_x^{(n)}(k+1, \cdot) = \gamma_x(y^k)$$

satisfies

$$\det A_x^{(n)} = 1.$$

Once again we will use induction. The base case is trivial to check. So suppose that we have our result for n. Proceeding as before, and noting that each of the first 2^n configurations begins with 0, so that $(x(y)_i, y_i) = (0,0)$ or $(1,0)$, and hence $\gamma(x(y)_i, y_i) = (1,0)$ or $(0,1)$, we see that there exists a function $\bar{x} : \Omega \to \Omega$ such that for each $0 \le k < 2^n$, either

$$\begin{aligned} A_x^{(n+1)}(k+1, \cdot) &= (1,0) \otimes \bigotimes_{i=2}^{n+1} \gamma_{\bar{x}}(y^k(i)) \\ &= \bigotimes_{i=2}^{n+1} \gamma_{\bar{x}}(y^k(i)) \oplus \bigoplus_{j=1}^{2^n} 0 \\ &= A_{\bar{x}}^{(n)}(k+1, \cdot) \oplus \bigoplus_{j=1}^{2^n} 0 \end{aligned}$$

or

$$\begin{aligned}
A_x^{(n+1)}(k+1,\cdot) &= (0,1) \otimes \bigotimes_{i=2}^{n+1} \gamma_{\bar{x}}(y^k(i)) \\
&= \bigoplus_{j=1}^{2^n} 0 \oplus \bigotimes_{i=2}^{n+1} \gamma_{\bar{x}}(y^k(i)). \\
&= \bigoplus_{j=1}^{2^n} 0 \oplus A_{\bar{x}}^{(n)}(k+1,\cdot)\ .
\end{aligned}$$

So, there exist $\varepsilon_k \in \{0,1\}$ such that

$$A_x^{(n+1)}(k+1,\cdot) = \varepsilon_{k+1} A_{\bar{x}}^{(n)}(k+1,\cdot) \oplus (1-\varepsilon_{k+1}) A_{\bar{x}}^{(n)}(k+1,\cdot).$$

Meanwhile, since $(1,-1) = (1,1)$ in $\mathbb{Z}_2^2$, there exists a function $\hat{x} : \Omega \to \Omega$ such that for each $2^n \le k < 2^{n+1}$

$$\begin{aligned}
A_x^{(n+1)}(k+1,\cdot) &= (1,1) \otimes \bigotimes_{i=2}^{n+1} \gamma_{\hat{x}}(y^k(i)) \\
&= \bigotimes_{i=2}^{n+1} \gamma_{\hat{x}}(y^k(i)) \oplus \bigotimes_{i=2}^{n+1} \gamma_{\hat{x}}(y^k(i)) \\
&= A_{\hat{x}}^{(n)}(k+1,\cdot) \oplus A_{\hat{x}}^{(n)}(k+1,\cdot)\ .
\end{aligned}$$

Hence

$$A_x^{(n+1)} = \begin{pmatrix} \varepsilon_k A_{\bar{x}}^{(n)}(k,j) & (1-\varepsilon_k) A_{\bar{x}}^{(n)}(k,j) \\ A_{\hat{x}}^{(n)} & A_{\hat{x}}^{(n)} \end{pmatrix}.$$

Although this matrix does not look quite as nice as the ones in the previous arguments, a few column operations—actually 2^n of them—will improve things, without changing the determinant, of course. By adding column $k+1$ to column $k+1+2^n$ (for each $0 \le k < 2^n$) which, in $\mathbb{Z}_2$, is the same as subtracting column $k+1$ from column $k+1+2^n$, we can conclude that

$$\begin{aligned}
\det A_x^{(n+1)} &= \det \begin{pmatrix} \varepsilon_k A_{\bar{x}}^{(n)}(k,j) & A_{\bar{x}}^{(n)}(k,\cdot) \\ A_{\hat{x}}^{(n)} & 0 \end{pmatrix} \\
&= \det A_{\bar{x}}^{(n)} \det A_{\hat{x}}^{(n)} \\
&= 1,
\end{aligned}$$

where the final step follows by induction. $\square$

The proof given above for independence of $\gamma(B)$ follows closely the matrix proof given by Reimer. Alternatively, there is a more algebraic proof, similar to one presented by J. Kahn and also to one suggested to us by P. Deligne. Below we give a version of such a proof, originally presented by one of us (J.T.C.) in the Kac Seminars of 1995.

Alternative Proof of Independence of $\gamma(B)$. As in the above proof, we will establish independence on $\mathbb{Z}_2^{2^n}$ rather than $\mathbb{R}^{2^n}$. For convenience, we define $\mathbf{w} = (1,0)$ and $\mathbf{v} = (1,1)$, so that $\gamma(0,0) = \mathbf{w}$, $\gamma(0,1) = \gamma(1,1) = \mathbf{v}$ and $\gamma(1,0) = \mathbf{w} + \mathbf{v}$. As in the proof above, we will use the fact that the tips are distinct to write $\gamma(x,y) = \gamma(x(y),y)$.

To show that $\gamma(x(y),y)$ are linearly independent, it suffices to expand them in a basis in $\otimes_{i=1}^n \mathbb{Z}_2^2$ and show that the coefficient matrix has nonzero determinant. To this end, let $I \subset \mathbb{B}$ be a label of a subset of sites. Our basis in $\otimes_{i=1}^n \mathbb{Z}_2^2$ will be $\{u_I \mid I \subset \mathbb{B}\}$ where

$$u_I = \bigotimes_{j \notin I} \mathbf{v} \otimes \bigotimes_{j \in I} \mathbf{w}\,.$$

In order to expand $\gamma(x(y),y)$ in the $\{u_I\}$, we let $I_y = \{i \mid y_i = 0\}$. Then

$$\begin{aligned}
\gamma(x(y),y) &= \bigotimes_{i=1}^n \gamma(x(y)_i, y_i) \\
&= \bigotimes_{i \notin I_y} \mathbf{v} \otimes \bigotimes_{i \in I} \begin{cases} \mathbf{w} & \text{if } x_i = 0 \\ \mathbf{w} + \mathbf{v} & \text{if } x_i = 1 \end{cases} \\
&= \bigotimes_{i \notin I_y} \mathbf{v} \otimes \bigotimes_{i \in I_y} (\mathbf{w} + x_i \mathbf{v}) \\
&= \sum_{J \subseteq I_y} \bigotimes_{i \notin I_y} \mathbf{v} \otimes \bigotimes_{i \in J} \mathbf{w} \otimes \bigotimes_{i \in I_y \setminus J} x_i \mathbf{v}\,.
\end{aligned}$$

Noting that $I_y^c \cup I_y \setminus J = J^c$, and defining $\epsilon(J) =: \prod_{I_y \setminus J} x_i \in \{0,1\}$, we have

$$\begin{aligned}
\gamma(x(y),y) &= \sum_{J \subseteq I_y} \epsilon(J) \bigotimes_{i \notin J} \mathbf{v} \otimes \bigotimes_{i \in J} \mathbf{w} \\
&= \sum_{J \subseteq I_y} \epsilon(J)\, u_J \\
&= u_{I_y} + \sum_{J \subset I_y} \epsilon(J)\, u_J\,,
\end{aligned}$$

where J is a proper subset of I_y in the final sum.

Now the above matrix is an upper triangular matrix with 1's along the diagonal. If the index set I_y were a totally ordered set, this would immediately imply $\det \gamma(x(y),y) = 1$ and hence that the $\gamma(x(y),y)$ are linearly independent. Since the index set is only partially ordered, this requires a little additional argument, which we leave to the reader. It is easy to verify using e.g. the expansion of the determinant in minors. □

CHAPTER VII

THE POTTS MODEL AND THE RANDOM CLUSTER MODEL

The Potts models are statistical mechanical models of magnets which turn out to be very closely related to percolation. Using the Potts models, we will be able to see precisely how P_∞ is analogous to the spontaneous magnetization and how $\tau(x,y)$ is analogous to the spin-spin correlation. The Potts models are also of interest in their own right. They have two parameters: β, an inverse temperature parameter, analogous to the density parameter p in percolation, and an additional parameter q, representing the number of possible states at each site. As we will see, the $q = 1$ model is equivalent to percolation, and the $q = 2$ model is equivalent to the Ising magnet. Both percolation and the Ising magnet have second-order, i.e. continuous, phase transitions in all dimensions. The $q = 3$ and $q = 4$ Potts models have second-order transitions in two dimensions, but first-order, i.e. discontinuous, phase transitions in higher dimensions. The $q \geq 5$ models have first-order transitions in all dimensions $d \geq 2$.

The Potts models are among the most widely studied of all statistical mechanics models. They are the simplest models to exhibit first-order transitions, and therefore are used by mathematical, theoretical and experimental physicists to model systems with such transitions. As we will discuss at length, they can be mapped into interacting percolation models [FK], and therefore are of interest to probabilists. Moreover, using this mapping, it is possible to simulate the Potts model with incredibly efficient algorithms [SW], and therefore the model is used a great deal by numerical physicists. For certain q, the Potts model is exactly (though not rigorously) solvable in two dimensions, and therefore is widely studied in the exact solutions community. Indeed, the model has fascinating combinatoric and algebraic content: The partition function of the Potts model is equivalent to the Tutte polynomial. It also arises in the fundamental equations of knot theory [W].

In this chapter, we will first define the Potts models in their conventional spin representation. Then, using a method due to Fortuin and Kasteleyn [FK], we will rewrite them as interacting percolation models. It will turn out that these interacting percolation models will be perfectly well defined even for q noninteger. The model with arbitrary real q is usually called the random cluster model.

We will review many of the properties of the random cluster model, including correlation inequalities, the nature of Gibbs states and the behavior of length scales in the model. In the process, we will cover many of the results of (but omit most of the proofs in) [ACCN2] and [BC]. Much of this chapter is taken almost verbatim from the paper of Borgs and Chayes [BC].

7.1 The Potts Models

Consider a finite site lattice $\Lambda \subset \mathbb{L}_d$. Let $S = \{1, \ldots, q\}$ denote the set of possible *spin states* at each site $x \in \Lambda$. The state space of the Potts model is $\Omega_\Lambda = S^\Lambda = \{1, \ldots, q\}^\Lambda$. A generic spin configuration $\sigma \in \Omega_\Lambda$ thus has spin values $\sigma_x \in S$ for each $x \in \Lambda$.

The *Hamiltonian* $H_\Lambda : \Omega_\Lambda \to \mathbb{R}$ is given by

$$H_\Lambda(\sigma) = H_\Lambda(\sigma; \{J_{x,y}\}) = -\sum_{(x,y)\in\Lambda\times\Lambda} J_{x,y}\left(\delta_{\sigma_x,\sigma_y} - 1\right) - h\sum_{x\in\Lambda}\delta_{\sigma_x,1} \tag{7.1}$$

where the *coupling* $J_{x,y}$ is a real symmetric function on $\mathbb{L}_d \times \mathbb{L}_d$, $\delta_{i,j}$ is the indicator that $i = j$, and the *external magnetic field* h is a real constant. In many instances, $J_{x,y}$ is taken to be translation invariant. In this case, $J_{x,y} = J(x - y)$ and, by the symmetry assumption, $J(y - x) = J(x - y)$. If $J_{x,y} \geq 0$ for all $(x, y) \in \mathbb{L}_d \times \mathbb{L}_d$, the system is said to be *ferromagnetic*. Often we only allow interaction between nearest-neighbor pairs which amounts to $J_{x,y} \equiv 0$ if $|x - y| > 1$.

Sometimes we take $\{J_{x,y}\}$ to be i.i.d. random variables with common distribution $\rho = \rho(J)$. Then the $J_{x,y}$ are clearly not translation invariant, although they are stationary in distribution. When the distribution ρ assigns substantial mass to both positive and negatives values of J, the system is a *spin glass*. When ρ has an atom at $J \equiv 0$, the system is said to be *dilute*. We will usually take $\{J_{x,y}\}$ to be fixed and suppress the $\{J_{x,y}\}$-dependence in our notation.

The first term in the Hamiltonian is the interaction term. It measures the amount of agreement between pairs of spins. This interaction is the most insensitive nontrivial trace on a group of q elements—it measures only whether spins agree or disagree, but not by how much. Note, however, if $q = 2$, this is the only nontrivial trace, and indeed the $q = 2$ Potts model is equivalent to the Ising magnet. For $q \geq 3$, there are other models besides the Potts model on the spin space $\Omega_\Lambda = S^\Lambda$. If J is nonnegative, the interaction favors alignment of the spins: the *energy* $H_\Lambda(\sigma)$ is lower if spins are aligned. The second term in $H_\Lambda(\sigma)$ is the external magnetic field term, here for a field in the 1-direction. It favors alignment with the field.

We define the *range* R of the Hamiltonian by

$$R = \sup_{x,y}\{|x - y| : J_{x,y} \neq 0\}.$$

Technically, we must include in the interaction term of H_Λ the sites in Λ^c within R of Λ. For this purpose, we let $\partial\Lambda_R$ denote the *boundary of* Λ (i.e., the set of sites $y \in \mathbb{L}_d \setminus \Lambda$ within distance R of Λ) and we replace the interaction term in (7.1) by

$$H_\Lambda\left(\{\sigma_x : x \in \Lambda\} \mid \{\sigma_y : y \in \partial\Lambda_R\}\right) = -\sum_{x\in\Lambda}\sum_{y\in\Lambda\cup\partial\Lambda_R} J_{x,y}\left(\delta_{\sigma_x,\sigma_y} - 1\right).$$

The set $\varphi = \{\sigma_y : y \in \partial\Lambda_R\}$ is called the *boundary condition*. Often we write $H_{\Lambda,\varphi}(\sigma)$ in place of $H_\Lambda(\cdot|\varphi)$. If we take $\varphi = \emptyset$, i.e. if we do not specify the spins on $\partial\Lambda_R$, then φ is said to be a free boundary condition. If we take $\sigma_y = c$ for all $y \in \partial\Lambda_R$, where c is a fixed element of $S = \{1, \dots q\}$, then φ is said to be the constant c boundary condition. Finally, if Λ is a d-dimensional torus so that $\partial\Lambda_R = \emptyset$, i.e. if there are no boundary sites, then the system is said to have periodic boundary conditions.

The *partition function* at inverse temperature β with boundary condition φ is defined by

$$Z_{\Lambda,\varphi}(\beta) = \sum_{\sigma \in \Omega_\Lambda} \exp\left(-\beta H_{\Lambda,\varphi}(\sigma)\right).$$

The partition function $Z_{\Lambda,\varphi}(\beta)$ induces a finite-volume probability measure on Ω_Λ given by

$$\nu_{\Lambda,\varphi}(\sigma) = \frac{\exp\left(-\beta H_{\Lambda,\varphi}(\sigma)\right)}{Z_{\Lambda,\varphi}(\beta)}.$$

For future reference, we note that in $d = 2$, it turns out that the partition function $Z_{\Lambda,\mathrm{free}}(\beta)$ is equal to the partition function $Z_{\Lambda,\mathrm{c}}(\beta^*)$ at the dual temperature β^* defined by

$$\left(e^{\beta} - 1\right)\left(e^{\beta^*} - 1\right) = q. \tag{7.2}$$

The self-dual point, i.e. the point at which $\beta = \beta^*$, is given by

$$\beta_{\mathrm{s}} = \log\left(1 + \sqrt{q}\right). \tag{7.3}$$

An *observable* A is a function (i.e., a random variable) on the space of configurations:

$$A : \Omega_\Lambda \to \mathbb{C}.$$

The expectation of A with respect to $\nu_{\Lambda,\varphi}$, denoted by $\langle A\rangle_{\Lambda,\varphi}$, is

$$\langle A\rangle_{\Lambda,\varphi} = \frac{1}{Z_{\Lambda,\varphi}(\beta)} \sum_{\sigma\in\Omega_\Lambda} A(\sigma) \exp\left(-\beta H_{\Lambda,\varphi}(\sigma)\right).$$

Using either the mapping to the random cluster model ([ACCN2], Theorem 2.3) or Griffith's second inequality [Grif], as generalized by Ginibre [Gin], it can be shown that for either free or constant boundary conditions, an infinite-volume Gibbs measure exists for any ferromagnetic Potts model. That is, that there exists a measure ν_φ such that

$$\nu_{\Lambda,\varphi} \to \nu_\varphi \tag{7.4}$$

as $\Lambda \nearrow \mathbb{L}_d$. This result does not require translation invariance or any restriction on the couplings $J_{x,y}$ other than nonnegativity.

Let us define the relevant quantities in the model. The *free energy* is the limit

$$f(\beta, h) = \lim_{\Lambda \nearrow \mathbb{L}_d} f_\Lambda(\beta, h)$$

where

$$f_\Lambda(\beta, h) = \frac{-\log\left(Z_{\Lambda,\varphi}(\beta)\right)}{|\Lambda|}.$$

Again under general conditions, the limit exists and is independent of boundary conditions. The *spontaneous magnetization* is given by

$$M(\beta) := \lim_{h \searrow 0} \frac{1}{\beta}\frac{\partial f}{\partial h} = \frac{q}{q-1}\left\langle\left(\delta_{\sigma_0,1} - \frac{1}{q}\right)\right\rangle_1 (\beta), \tag{7.5}$$

where $\langle\cdot\rangle_1$ is the thermal average in the (infinite-volume) Gibbs measure constructed with the constant $\varphi_y \equiv 1$ boundary condition. Finally, the *two-point correlation function* with boundary condition φ is

$$\left\langle q\delta_{\sigma_x,1}; q\delta_{\sigma_y,1}\right\rangle_\varphi (\beta) = \frac{q}{q-1}\left\langle\left(\delta_{\sigma_x,\sigma_y} - \frac{1}{q}\right)\right\rangle_\varphi (\beta).$$

As we will see, the magnetization is the analogue of the percolation probability P_∞ and the two-point function is the analogue of the connectivity function $\tau(x,y)$.

In order to obtain alternate expressions for the magnetization and the two-point, or spin-spin, correlation function, it is sometimes useful to represent the spin variables differently. Rather than $\sigma_x \in S$, we can take the spin variables to be unit vectors $\vec{\sigma}_x \in \mathbb{R}^{q-1}$ from the center to the (equally-spaced) corners of a $(q-1)$-dimensional tetrahedron.

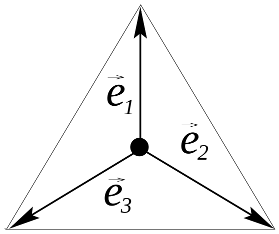


FIGURE 7.1. *The spin vectors for $q = 3$. Here, the $(q-1)$-dimensional tetrahedron is simply an equilateral triangle.*

Note that

$$\vec{\sigma}_x \cdot \vec{\sigma}_y = \begin{cases} 1 & \text{if } \vec{\sigma}_x = \vec{\sigma}_y, \\ -\frac{1}{q-1} & \text{otherwise.} \end{cases}$$

Hence, we can write

$$\vec{\sigma}_x \cdot \vec{\sigma}_y = \frac{q\delta_{\sigma_x,\sigma_y} - 1}{q-1},$$

and, up to a constant, the Hamiltonian can be rewritten as

$$H(\sigma) = -\sum_{x,y} \tilde{J}_{x,y}(\vec{\sigma}_x \cdot \vec{\sigma}_y - 1) - \tilde{h}\sum_x \vec{e}_1 \cdot \vec{\sigma}_x$$

where $\tilde{J}_{x,y} = \frac{q-1}{q}J_{x,y}$, $\tilde{h}_{x,y} = \frac{q-1}{q}h$ and $\vec{e}_1$ is the unit vector in the spin 1-direction. In this alternate representation, the magnetization is $M(\beta) = \langle\vec{e}_1 \cdot \vec{\sigma}_0\rangle_1(\beta)$, and the two-point function is simply $\langle\vec{\sigma}_x \cdot \vec{\sigma}_y\rangle_\varphi(\beta)$.

Let us discuss the expected behaviors of M and $\langle \vec{\sigma}_x \cdot \vec{\sigma}_y \rangle$ for various values of q. The magnetization is the order parameter of the Potts phase transition. Noting that $M(\beta)$ is increasing in β, we see that it defines the transition point

$$\beta_{\mathrm{t}} = \inf\{\beta : M(\beta) > 0\}. \tag{7.6}$$

We remark that in $d = 2$, β_{t} is expected, though not proven, to coincide with the self-dual point β_{s} defined in equation (7.3). As mentioned at the beginning of this section, the Potts model is believed to have a second-order, i.e. continuous, transition for values of q up to some critical value, and to have a first-order, i.e. discontinuous, transition above this value. Assuming that $M(\beta_{\mathrm{t}})$ is increasing in q, this value may be defined by

$$q_{\mathrm{c}} = q_{\mathrm{c}}(d) = \max\{q \in \mathbb{N} : M(\beta_{\mathrm{t}}) = 0\}.$$

FIGURE 7.2. *The expected behaviors of $M(\beta)$ in the cases $q < q_{\mathrm{c}}$ and $q > q_{\mathrm{c}}$, respectively.*

It is believed that $q_{\mathrm{c}} = 4$ in $d = 2$ and $q_{\mathrm{c}} = 2$ in $d > 2$. The fact that $M(\beta_{\mathrm{t}}) > 0$ has been established rigorously using expansion methods for all $d \geq 2$ provided q is sufficiently large ([KS], [LMR], [LMMRS]).

It is expected that the two-point function with free boundary conditions decays exponentially for $\beta < \beta_{\mathrm{t}}$:

$$\langle \vec{\sigma}_x \cdot \vec{\sigma}_y \rangle_{\text{free}} \sim e^{-|x-y|/\xi(\beta)},$$

where, as usual, $\sim$ means that the limit of the ratio of the logarithms with x and y along a coordinate axis. Here, as in percolation, we identify ξ as the correlation length of the system. It is also expected that $\xi(\beta) \nearrow \infty$ as $\beta \nearrow \beta_{\mathrm{t}}$ when $q < q_{\mathrm{c}}$, while $\xi(\beta_{\mathrm{t}}) < \infty$ for $q > q_{\mathrm{c}}$, as illustrated below.

FIGURE 7.3. *The expected behaviors of $\xi(\beta)$ in the cases $q < q_c$ and $q > q_c$, respectively.*

As we have just seen, models with first- and second-order transitions are typically distinguished by the continuity properties of relevant quantities such as the correlation length. It turns out that there is another way to distinguish them: by the number of states at the transition point. For $\beta < \beta_t$, all of the Potts models have 1 (pure) state, i.e. 1 ergodic component of any translation-invariant measure. For $\beta > \beta_t$, the q-state Potts model has q pure states; they may be induced by applying $\sigma_x \equiv s \in S$ boundary conditions. However, the number of states at $\beta = \beta_t$ depends on the order of the transition. For $q < q_c(d)$, the q-state model has only 1 pure state at β_t. This should be contrasted with the $q > q_c(d)$ models which have $q+1$ pure states, and hence phase coexistence of the high- and low-temperature phases, at β_t. The phases are separated by interfaces with a positive surface tension, which is of course related to the finite correlation length discussed above. Phase coexistence has been rigorously established for q sufficiently large [LMR].

7.2 The Fortuin-Kasteleyn Representation

Fortuin and Kasteleyn [FK] re-expressed the q-state Potts model as an integer value of a two-parameter interacting percolation model, the so-called random cluster model. Given a set of couplings $\{J_{x,y}\}$, the representation is defined in terms of configurations $\omega \in \Omega \equiv \{0,1\}^{\mathbb{K}_d}$, where

$$\mathbb{K}_d = \mathbb{K}_d(\{J_{x,y}\}) = \{\langle x,y\rangle : x, y \in \mathbb{L}_d \quad J_{x,y} \neq 0\}$$

is the bond lattice on $\mathbb{L}_d$ consisting of bonds between all pairs of points with nonzero interaction. For subsets $K \subset \mathbb{K}_d$, the configuration space is denoted by $\Omega_K = \{0,1\}^K$. Given a finite subset $\Lambda \subset \mathbb{L}_d$, let $K(\Lambda)$ denote the corresponding $\{J_{x,y}\}$-dependent bond graph on Λ.

Throughout the rest of this chapter, we will restrict attention to the Potts model with $h = 0$, but we note that the FK representation can also be derived for nonzero external magnetic field.

We begin by obtaining the representation for the finite-volume partition function with free boundary conditions. We write the Gibbs factor $e^{-\beta H_{\Lambda,\mathrm{free}}(\sigma)}$ as

$$\prod_{(x,y)\in K(\Lambda)} e^{\beta J_{x,y}(\delta_{\sigma_x,\sigma_y}-1)},$$

and expand the product with the help of the identity

$$e^{\beta J_{x,y}(\delta_{\sigma_x,\sigma_y}-1)} = (1-p_{x,y}) + p_{x,y}\delta_{\sigma_x,\sigma_y} \quad \text{where} \quad p_{x,y} = 1 - e^{-\beta J_{x,y}}.$$

Each term of this expansion can be identified with a configuration $\omega \in \Omega_{K(\Lambda)}$: ω is chosen so that it is zero on those bonds for which the factor in the product is $1 - p_{x,y}$, and one on those bonds for which the factor is $p_{x,y}\delta_{\sigma_x,\sigma_y}$. Geometrically, we think of the bonds $b=(x,y)$ for which $\omega_b=1$ as occupied, and those for which $\omega_b=0$ as vacant. Sometimes we use the symbol ω to denote the set of occupied bonds in $K(\Lambda)$, and ω^c to denote the set of vacant bonds in $K(\Lambda)$.

Rewriting the Gibbs factor in expanded form, we obtain

$$\begin{aligned} Z_{\Lambda,\mathrm{free}}(\beta) &= \sum_{\sigma\in\Omega_\Lambda} \sum_{\omega\in\Omega_{K(\Lambda)}} \prod_{\{b:\omega_b=0\}} (1-p_b) \prod_{\{b:\omega_b=1\}} p_b \prod_{\{(x,y)=b:\omega_b=1\}} \delta_{\sigma_x,\sigma_y} \\ &= \sum_{\omega\in\Omega_{K(\Lambda)}} \prod_{\{b:\omega_b=0\}} (1-p_b) \prod_{\{b:\omega_b=1\}} p_b \sum_{\sigma\in\Omega_\Lambda} \prod_{\{(x,y)=b:\omega_b=1\}} \delta_{\sigma_x,\sigma_y} \,. \end{aligned}$$

All that remains to do is evaluate the sum over $\sigma\in\Omega_\Lambda$. As usual, ω partitions Λ into connected components or clusters. The delta functions require that all spins within a given component assume the same value—any of the q possible spins values. Otherwise, the trace is unconstrained. Thus the trace gives a factor of q for each connected component of the graph (Λ,ω) (regarding isolated points as separate clusters). Denoting the number of clusters in this graph by $\#(\omega)$, we find

$$Z_{\Lambda,\mathrm{free}}(\beta) = \sum_{\omega\in\Omega_{K(\Lambda)}} \prod_{\{b:\omega_b=0\}} (1-p_b) \prod_{\{b:\omega_b=1\}} p_b \; q^{\#(\omega)} \,.$$

Letting $\mathcal{B}_p(\omega)=\prod_{\{b:\omega_b=0\}}(1-p_b)\prod_{\{b:\omega_b=1\}}p_b$ denote the Bernoulli weight of ω with bond values $p=\{p_b\}=\{1-e^{-\beta J_b}\}$, we finally obtain

$$Z_{\Lambda,\mathrm{free}}(\beta) = \sum_{\omega\in\Omega_{K(\Lambda)}} \mathcal{B}_p(\omega) q^{\#(\omega)} \,.$$

Similarly, we may write the expectation of of any local observable A as

$$\langle A\rangle_{\Lambda,\mathrm{free}} = \sum_{\omega\in\Omega_{K(\Lambda)}} G_{\Lambda,\mathrm{free}}(\omega) E_{\mathrm{free}}(A|\omega) \;, \tag{7.7}$$

where

$$G_{\Lambda,\mathrm{free}}(\omega) = \frac{\mathcal{B}_p(\omega) q^{\#(\omega)}}{Z_{\Lambda,\mathrm{free}}(\beta)} \tag{7.8}$$

is the weight of the configuration ω, while $E_{\mathrm{free}}(\cdot|\omega)$ is an average over spins with the spins constrained to be constant on each connected cluster of ω and with values for different clusters being chosen uniformly from $\{1,\cdots,q\}$. Note that free boundary conditions in the spin representation transform into free boundary conditions in the FK representation.

For constant boundary conditions, one obtains a similar representation, with the following differences (as noted in [ACCN2] and [BC]):

(i) The set $K(\Lambda)$ is replaced by the set $K^+(\Lambda)$ of all pairs $\langle x,y\rangle$ for which at least one of the two points x and y lies in Λ and the other lies in $\Lambda\cup\partial\Lambda_R$.

(ii) The points of the boundary $\partial\Lambda_R$ are regarded as *preconnected* or *wired*, in the sense that these points are taken to be lying in one cluster. This of course decreases the value of $\#(\omega)$ by one less than the number of connected components of Λ connecting to $\partial\Lambda_R$.

(iii) The expectation $E_{\text{free}}(A|\omega)$ in is replaced by $E_c(A|\omega)$, where the average is computed with the additional constraint that spins in clusters connected to the boundary now assume only the value $\sigma_x = c$. Note that all q constant boundary condition measures in the spin representation transform into a single measure in the FK representation. We call this the *wired measure.*

We have:

$$\langle A\rangle_{\Lambda,c} = \sum_{\omega\in\Omega_{K^+(\Lambda)}} G_{\Lambda,\text{wir}}(\omega)E_c(A|\omega)\,, \tag{7.9}$$

where

$$G_{\Lambda,\text{wir}}(\omega) = \frac{\mathcal{B}(p)q^{\#(\omega)}}{Z_{\Lambda,\text{wir}}(\beta)} \tag{7.10}$$

and $Z_{\Lambda,\text{wir}} = \sum_{c\in S} Z_{\Lambda,c} = qZ_{\Lambda,1}$.

We denote by $\mu_{\Lambda,\text{free}}(\cdot)$ and $\mu_{\Lambda,\text{wir}}(\cdot)$ the finite-volume measures defined by the weights (7.8) and (7.10), respectively, and we denote by $\mu_{\Lambda,\text{per}}(\cdot)$ the analogously defined measure with periodic boundary conditions. We can of course derive finite-volume measures with other boundary conditions. For example, in the spin system we could take so-called Dobrushin boundary conditions in which the spins in the upper half of $\partial\Lambda_R$ are taken to be in the $\vec{e}_1$ direction, while those in the bottom half of $\partial\Lambda_R$ are taken to be in the $\vec{e}_2$ direction. This would mean that the resulting FK measure gives zero weight to configurations ω with components connecting the upper and lower halves of $\partial\Lambda_R$. The problem with such a finite-volume measure is that, at present, we do not know how to extract an infinite-volume limit. On the other hand, we can establish convergence of the measures $\mu_{\Lambda,\text{free}}(\cdot)$, $\mu_{\Lambda,\text{wir}}(\cdot)$, and $\mu_{\Lambda,\text{per}}(\cdot)$. Indeed (again by [ACCN2], Theorem 2.3), for free, wired, or periodic boundary conditions, $p\in[0,1]$, and $q\geq 1$, there exists a measure μ_φ such that

$$\mu_{\Lambda,\varphi}\to\mu_\varphi \tag{7.11}$$

as $\Lambda\nearrow\mathbb{L}_d$. The proof relies on the FKG inequality, which we will establish for the random clusters measures in the next subsection.

Notice that the random cluster model is a *dependent percolation* model. It has percolation weights $\mathcal{B}_p(\omega)$ modified by the factor $q^{\#(\omega)}$. Not surprisingly, the magnetization and spin-spin correlation have expressions which look like percolation probabilities and connectivities. Indeed, using monotonicity and convexity properties, one can show ([ACCN2], Lemma 2.1 and Theorem 2.3) that

$$M(\beta) \;=\; \mu_{\text{wir}}(0\leftrightarrow\infty)\,. \tag{7.12}$$

A much more involved proof ([BC], Proposition 3.4) shows also that

$$\langle\vec{\sigma}_x\cdot\vec{\sigma}_y\rangle_{\text{free}} \;=\; \mu_{\text{free}}(x\leftrightarrow y) \;:=\; \tau_{\text{free}}(x,y). \tag{7.13}$$

The random cluster representation is often more useful than the spin representation, but both have their attributes. Edwards and Sokal [ES] pointed out that there is a joint distribution $\lambda_{\Lambda,\varphi}$ on $\Omega_{K(\Lambda)} \times \Omega_\Lambda$ which has both the FK measure $\mu_{\Lambda,\varphi}$ and the spin measure $\nu_{\Lambda,\varphi}$ as its marginals:

$$\lambda_{\Lambda,\varphi}(\sigma,\omega) \propto \prod_{(x,y)\in K(\Lambda)} \left((1-p_{x,y})\delta_{\omega_{x,y},0} + \delta_{\sigma_x,\sigma_y} p_{x,y} \delta_{\omega_{x,y},1}\right).$$

Given these weights, it is clear that:

(i) The conditional distribution of ω given σ can be obtained as follows: For each bond $b = \langle x,y\rangle$,
  (a) if $\sigma_x \neq \sigma_y$, then set $\omega_{x,y} = 0$, while
  (b) if $\sigma_x = \sigma_y$, then set

$$\omega_{x,y} = \begin{cases} 0 & \text{with probability } 1-p_{x,y} \\ 1 & \text{with probability } p_{x,y}. \end{cases}$$

(ii) The conditional distribution of σ given ω can be obtained as follows: Given a cluster $\mathcal{C}$ (potentially an isolated site) of ω, choose uniformly over the q possible spin values and set all spins in $\mathcal{C}$ to that value.

There is a very efficient computer algorithm, called the *Swendsen-Wang* algorithm [SW], which alternatingly applies (i) and (ii) to approximate λ. There is no rigorous argument which shows that this procedure actually converges to the correct infinite-volume distribution. However, the algorithm is widely believed to yield accurate results, and is in practice much faster than the conventional Monte Carlo algorithms based on the spin representation.

7.3 Standard Correlation Inequalities

7.3.1 FKG Inequality.

An important property of the FK representation is that, for $q \geq 1$, it obeys the Harris-FKG inequality, as discussed for percolation in Section 1.2.2. That the $q \geq 1$ random cluster measure satisfies the FKG inequality is a consequence of the following very useful proposition.

PROPOSITION 7.1. ([FKG]) *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space with μ a measure of the form $d\mu = f d\rho$ where ρ is a product measure and f is a nonnegative function satisfying*

$$f(\omega \vee \omega')f(\omega \wedge \omega') \geq f(\omega)f(\omega') \quad \forall \omega, \omega' \in \Omega\,. \tag{7.14}$$

Here, as usual,

$$(\omega \vee \omega')_b = \max(\omega_b, \omega'_b)$$
$$(\omega \wedge \omega')_b = \min(\omega_b, \omega'_b).$$

Then μ is an FKG measure; that is, for all increasing A, $B \in \mathcal{F}_\Lambda$,

$$\mu(A \cap B) \geq \mu(A)\mu(B).$$

We will not bother to prove this here, referring the reader instead to [FKG]. Note, however, that we have already seen the proof in case $f \equiv 1$ in Theorem 1.3. Measures which satisfy the sufficient condition in Proposition 7.1 are said to be *strong FKG* measures.

THEOREM 7.2. ([Fo], [ACCN2]) *Let $q \geq 1$. Then the finite-volume free, periodic, and wired FK measures, $\mu_{\text{free},\Lambda}$, $\mu_{\text{per},\Lambda}$, and $\mu_{\text{wir},\Lambda}$, are strong FKG measures.*

PROOF. The random cluster measures with free or periodic boundary conditions are explicitly of the form specified in the FKG theorem, while the measure with wired boundary conditions is equivalent to a free problem on a lattice in which $\partial\Lambda_R$ has been collapsed to a single point. Thus it suffices to show that the function

$$f(\omega) = q^{\#(\omega)}$$

satisfies (7.14) for all ω, $\omega' \in \Omega_{K(\Lambda)}$. For $q \geq 1$, (7.14) is equivalent to

$$\#(\omega \vee \omega') \; + \; \#(\omega \wedge \omega') \geq \#(\omega) + \#(\omega'). \tag{7.15}$$

Suppose we show that for all ω, η, $\zeta \in \Omega_{K(\Lambda)}$ such that $\eta \succeq \zeta$,

$$\#(\omega \vee \eta) - \#(\eta) \geq \#(\omega \vee \zeta) - \#(\zeta). \tag{7.16}$$

Then given ω, $\omega' \in \Omega_{K(\Lambda)}$, let $\eta = \omega'$ and $\zeta = \omega \wedge \omega'$ and note that $\omega \vee \zeta = \omega \vee (\omega \wedge \omega') = \omega$. Hence, (7.16) implies (7.15). Finally, (7.16) is implied by the statement that

$$g(\omega, \eta) \; = \; \#(\omega \vee \eta) - \#(\eta) \tag{7.17}$$

is an increasing function of η. We prove this by induction on the number of occupied bonds in ω.

The base case holds trivially since when ω is a single bond b, we have

$$g(\omega, \eta) \; = \; \begin{cases} -1 & \text{if the endpoints of } b \text{ are not connected in } \eta, \\ 0 & \text{otherwise.} \end{cases}$$

Assume that $g(\omega, \eta)$ is an increasing function of η for ω with less than n occupied bonds. Given ω with exactly n bonds occupied, we can write $\omega = \omega_1 \vee \omega_2$ where ω_1 has exactly one bond occupied and ω_2 has exactly $n-1$ bonds occupied. Using the definition of $g(\omega, \cdot)$ and adding and substracting the term $\#(\omega_2 \vee \eta)$, we have

$$\begin{aligned} g(\omega, \eta) &= \; g(\omega_1 \vee \omega_2, \eta) \\ &= \; g(\omega_1, \omega_2 \vee \eta) + g(\omega_2, \eta). \end{aligned}$$

Noting that $\zeta \succeq \eta$ implies $\omega_2 \vee \zeta \succeq \omega_2 \vee \eta$ gives that $g(\omega_1, \omega_2 \vee \cdot)$ is an increasing function. Hence, $g(\omega, \cdot)$ is the sum of two increasing functions and is therefore increasing. $\square$

We have just shown that the $q \geq 1$ random cluster measures are FKG. What about the $q < 1$ random cluster measures? We have the following:

COUNTEREXAMPLE: Fix a bond b such that $0 < p_b < 1$. Then

$$\mu_\Lambda(\omega_b = 1 \,|\, \omega_{b'}, b' \neq b) = \begin{cases} p_b & \text{if the endpoints of } b \text{ are connected, regardless of } \omega_b \\ \frac{p_b}{p_b + q(1-p_b)} & \text{otherwise.} \end{cases} \tag{7.18}$$

Notice that

$$\frac{p_b}{p_b + q(1-p_b)} \begin{cases} > p_b & \text{if } q < 1, \\ = p_b & \text{if } q = 1, \\ < p_b & \text{if } q > 1. \end{cases}$$

Take A to be the event that b' is occupied for all $b' \neq b$ and let B be the event that b is occupied. Then

$$\mu_\Lambda(B|A) = p_b\,,$$

while

$$\begin{aligned} \mu_\Lambda(B) &= \sum_{\{\omega_{b'}\}_{b'\neq b}} \mu_\Lambda(\omega_b = 1 | \{\omega_{b'}\}_{b'\neq b})\ \mu_\Lambda(\{\omega_{b'}\}_{b'\neq b}) \\ &= \alpha p_b + (1-\alpha)\frac{p_b}{p_b + q(1-p_b)} \\ &\begin{cases} > p_b & \text{if } q < 1, \\ = p_b & \text{if } q = 1, \\ < p_b & \text{if } q > 1. \end{cases} \end{aligned} \tag{7.19}$$

Thus $q < 1$ implies

$$\mu_\Lambda(B) > \mu_\Lambda(B|A) = \frac{\mu_\Lambda(A \cap B)}{\mu_\Lambda(A)}\,,$$

so that FKG fails.

7.3.2 Failure of the BK Inequality.

The BK inequality is in some sense complementary to the FKG inequality, and therefore one might expect it to fail for the $q > 1$ random cluster measures. Indeed, our FKG counterexample described above gives a counterexample here as well.

Counterexample 1: Since the events A and B specified above are disjoint,

$$A \cap B = A \circ B.$$

However, for $q > 1$, (7.19) implies that

$$\mu_\Lambda(B) < \mu_\Lambda(B|A) = \frac{\mu_\Lambda(A \cap B)}{\mu_\Lambda(A)},$$

which violates the BK inequality.

Counterexample 2: It is also possible to give a more physical counterexample which shows that the failure of the BK inequality is intimately associated with the discontinuous nature of the Potts' phase transition. Let q be sufficiently large to guarantee that the Potts model has a first-order phase transition. Then it turns out that any translation-invariant random cluster measure at $p_c(q)$ will decompose into two ergodic components, one representing the high-density and the other the low-density phase–i.e., one with an infinite cluster and the other without an infinite cluster). Now consider the event of two disjoint crossings:

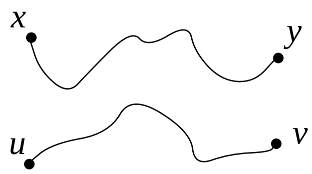


FIGURE 7.4. *Two disjoint crossings.*

If we know that the top crossing has occurred, then we know that the system is more likely to be in the ergodic component containing an infinite cluster. In this case, the bottom crossing is more likely to occur, even though it must do so disjointly from the top crossing. Again, the BK inequality is violated.

In the next section, we will see that even though the BK inequality fails for the $q > 1$ random cluster measures, we can in certain cases get a result analogous to the BK inequality by using the decoupling inequalities of Borgs and Chayes [BC].

It is an open, and very interesting question, whether the BK inequality holds for increasing events in the $q < 1$ random cluster measures. We know that the BK inequality cannot hold for *all* events in the $q < 1$ random cluster measures, since e.g. the BK inequality coincides the FKG inequality for the intersection of an increasing and a decreasing event, and the latter does not hold for $q < 1$. However, if the BK inequality held for increasing events in the $q < 1$ (finite-volume) random cluster measures, this could be used to show the existence of the corresponding infinite-volume measures.

7.4 The BC Decoupling Inequalities

There are three standard technical tools for factoring intersections of events in independent percolation: the FKG inequality for monotone increasing or decreasing events, independence for events which occur on nonrandom disjoint sets, and the BK inequality for events which occur on random disjoint sets. As discussed in the last subsection, the $q \geq 1$ free and wired random cluster measures obey an FKG inequality. However, due to the nonlocality of the weights (7.8) and (7.10), they do not satisfy a BK inequality. Nor of course do they satisfy an independence relation. Here we will review the BC decoupling inequalities [BC], which are alternatives to independence and the BK inequality for many events of interest in a general setting.

As a substitute for independence of events occuring on nonrandom disjoint sets, we might try to use the FKG inequality as a bound, provided that the desired events are monotone. However, many of the events we might want to consider—especially in the low-temperature phase—are not monotone. For example, the probability of a connection via finite clusters is the *intersection* of an increasing and a decreasing event, and the FKG inequality does not hold for two events of this form. The presence of boundary conditions, which very often complicates proofs in the random cluster model, can be used to our advantage here. Certain boundary conditions *decouple* a set from its exterior. Many events of interest carry with them decoupling boundary conditions for the (random) sets on which they occur. [BC] make this notion precise by introducing the definition of a *decoupling event*, see below. It turns out that, given this definition, it is possible to prove a general inequality which is similar to the FKG inequality and which replaces independence for events whose random boundaries occur within disjoint nonrandom sets. This inequality holds for any FKG measure and for events which are intersections of *arbitrary* events with monotone decoupling events.

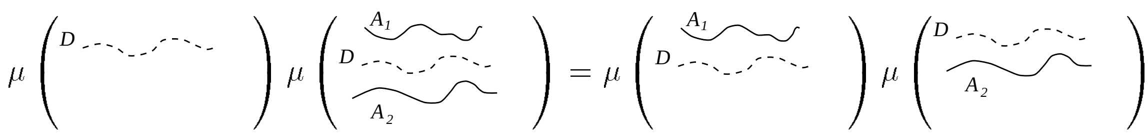
As illustrated above, the BK inequality is certainly not true in general for the random cluster model—there are numerous examples in which the occurrence of one event enhances the occurrence of another. However, this enhancement cannot take place if the two events are *decoupled* from one another, in a sense to be made precise in the definition below. Thus [BC] prove a second inequality, which replaces the BK inequality of independent percolation, and which holds for the intersection of an arbitrary event, an increasing event and a decreasing decoupling event.

In the next proposition, we actually present two versions of each of the BC inequalities: one which is easy to formulate (but not that useful), and a more involved one which is of the form needed in most applications. The more involved forms are simply generalizations of the simple forms to countable disjoint unions. All of these inequalities hold for general FKG measures. In particular, they hold for percolation, for the general $q \geq 1$ random cluster model, and for the Ising and Potts models in the spin representation. We begin with the definition of a decoupling event.

DEFINITION: Given a probability space $(\Omega, \mathcal{F}, \mu)$ and events $A_1, A_2, D \in \mathcal{F}$, we say that D is a *decoupling event* for A_1 and A_2, if

$$\mu(A_1 \cap A_2 \mid D) = \mu(A_1 \mid D)\, \mu(A_2 \mid D)\,, \tag{7.20}$$

that is, if A_1 and A_2 are conditionally independent, conditioned on D. For brevity, we will sometimes say D *decouples* A_1 from A_2. Pictorially, we write



While this definition makes sense in any probability space, it may be useful to illustrate it with a typical example from the nearest-neighbor uniform random cluster model. Consider a set $B \subset \mathbb{B}_d$ which divides the lattice into two components, an interior and an exterior: $\mathbb{B}_d \setminus B = B_1 \cup B_2$, $B_1 \cap B_2 = \emptyset$. The event that the bonds of B are vacant then decouples any event on the interior $A_1 \in \mathcal{F}_{B_1 \cup B}$ from any event on the exterior $A_2 \in \mathcal{F}_{B_2 \cup B}$. Such decoupling events typically occur when B is the boundary of a finite occupied cluster. Returning to the general context of a decoupling event, we have:

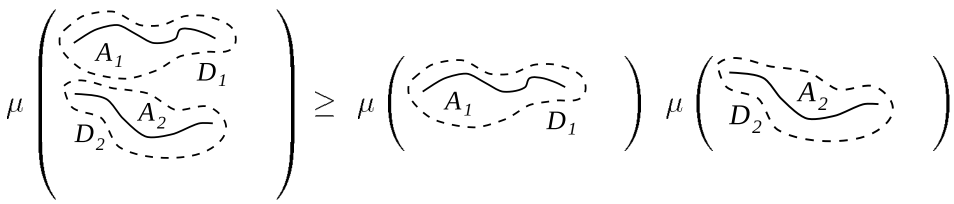
PROPOSITION 7.3. ([BC]) *Let $(\Omega, \mathcal{F}, \mu)$ be a probability space with Ω partially ordered and μ an FKG measure with respect to this order. Then the following inequalities hold:*

The Simple Versions

(i) *Consider two arbitrary events $A_1, A_2 \in \mathcal{F}$, and two increasing (or two decreasing) events $D_1, D_2 \in \mathcal{F}$ such that D_1 decouples A_1 from D_2 while D_2 decouples A_2 from $A_1 \cap D_1$. Then $E_1 = A_1 \cap D_1$ and $E_2 = A_2 \cap D_2$ obey the inequality*

$$\mu(E_1 \cap E_2) \geq \mu(E_1)\,\mu(E_2), \tag{7.21}$$

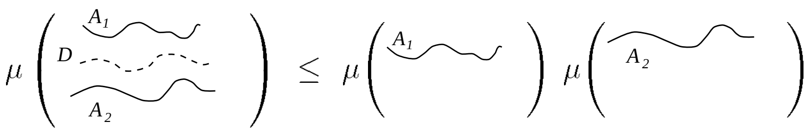
or pictorially



(ii) *Let $A_1 \in \mathcal{F}$ be an increasing event, $A_2 \in \mathcal{F}$ be arbitrary, and $D \in \mathcal{F}$ be a decreasing event which decouples A_1 from A_2. Then*

$$\mu(A_1 \cap D \cap A_2) \leq \mu(A_1)\mu(D \cap A_2) \leq \mu(A_1)\mu(A_2), \tag{7.22}$$

or pictorially



The Disjoint Union Versions

(i) *More generally, let E_i, $i = 1, 2$, be disjoint unions of the form*

$$E_i = \bigcup_{n \in N_i} A_{i,n} \cap D_{i,n}, \tag{7.23}$$

where N_i are countable index sets, $A_{i,n} \in \mathcal{F}$ are arbitrary events, $D_{i,n} \in \mathcal{F}$ are all increasing (or all decreasing) events, and $D_{1,n}$ decouples $A_{1,n}$ from $D_{2,n'}$ while $D_{2,n'}$ decouples $A_{2,n'}$ from $A_{1,n} \cap D_{1,n}$ for all $n \in N_1$ and $n' \in N_2$. Then E_1 and E_2 obey the inequality (7.21).

(ii) *Let $A_1 \in \mathcal{F}$ be an increasing event, and let $A_2 \in \mathcal{F}$ and $D \in \mathcal{F}$ be events for which $D \cap A_2$ can be rewritten as a disjoint union of the form (7.23), with $D_{2,n}$ decreasing events that decouple A_1 from $A_{2,n}$ for all $n \in N_2$. Then the bound (7.22) remains valid.*

PROOF. Rewriting the left hand side of (7.21) as

$$\mu(D_2)\,\mu(A_1 \cap D_1 \cap A_2 \mid D_2)$$

and using the fact that D_2 decouples A_2 from $A_1 \cap D_1$, we obtain

$$\mu(A_1 \cap D_1 \cap A_2 \cap D_2) = \mu(A_1 \cap D_1 \cap D_2)\,\mu(A_2 \mid D_2).$$

Applying the same procedure to the term $\mu(A_1 \cap D_1 \cap D_2)$ and using the decoupling event D_1, we get

$$\begin{aligned}
\mu(A_1 \cap D_1 \cap A_2 \cap D_2) &= \mu(A_1 \cap D_1 \cap D_2)\,\mu(A_2 \mid D_2) \\
&= \mu(A_1 \cap D_2 \mid D_1)\,\mu(D_1)\,\mu(A_2 \mid D_2) \\
&= \mu(A_1 \mid D_1)\,\mu(D_2 \mid D_1)\,\mu(D_1)\,\mu(A_2 \mid D_2) \\
&= \mu(A_1 \mid D_1)\,\mu(A_2 \mid D_2)\,\mu(D_1 \cap D_2)
\end{aligned}$$

By the FKG inequality,

$$\begin{aligned}
\mu(D_1 \cap D_2)\,\mu(A_1 \mid D_1)\,\mu(A_2 \mid D_2) &\geq \mu(D_1)\,\mu(D_2)\,\mu(A_1 \mid D_1)\,\mu(A_2 \mid D_2) \\
&= \mu(A_1 \cap D_1)\,\mu(A_2 \cap D_2)
\end{aligned}$$

proving the simple version of (7.21). The more general version of the proposition then follows from the countable additivity of the measure μ and the fact that the events E_1 and E_2 are disjoint unions of events for which (7.21) is valid.

In order to prove the simple version of (7.22), we observe that

$$\mu(A_1 \cap D \cap A_2) = \mu(D)\,\mu(A_1 \mid D)\,\mu(A_2 \mid D)$$

by the definition of conditional expectations and a decoupling event. Using the FKG inequality, we have

$$\begin{aligned}
\mu(D)\,\mu(A_1 \mid D)\,\mu(A_2 \mid D) &= \mu(A_1 \cap D)\,\mu(A_2 \mid D) \\
&\leq \mu(D)\,\mu(A_1)\,\mu(A_2 \mid D) \\
&= \mu(A_1)\,\mu(A_2 \cap D) \\
&\leq \mu(A_1)\,\mu(A_2)
\end{aligned}$$

which proves (7.22) in the simple case. Again the more general version follows from the simple version of (7.22) and the countable additivity of the measure. □

REMARK. It is clear from the above proof that the inequality (7.21) is reversed if one of the two decoupling events D_1 and D_2 is increasing and the other is decreasing. Similarly, the first inequality in (7.22) is reversed if A_1 and D are both decreasing or both increasing.

So far, there have been two types of applications of the BC decoupling inequalities. One type is technical, but quite useful: The inequalities allow us to establish convergence of certain quantities in the random cluster model, for example convergence of the so-called finite-volume connectivity function ([BC], Proposition 3.4):

$$\begin{aligned}
\tau_{\text{wir}}^{\text{fin}} &= \mu_{\text{wir}}(x \leftrightarrow y, |\mathcal{C}(x)| < \infty) \\
&= \lim_{\Lambda \nearrow \mathbb{L}_d} \mu_{\text{wir},\Lambda}(x \leftrightarrow y\ ,\ x \not\leftrightarrow \partial\Lambda).
\end{aligned} \tag{7.24}$$

The observant reader might at first assume that this convergence would follow immediately from the existence of the infinite-volume limiting measure (7.11). However, the problem here is the nonlocality of the event $\{x \leftrightarrow y\}$—the event may not be contained in any given bounded region. This leads to an interchange of limits: namely, the limit defining the event $\{x \leftrightarrow y\}$ and the limit defining the measure $\mu_{\rm wir}$. The interchange is not entirely technical; it is related to the way the infinite cluster emerges from large clusters in a finite volume. In fact, the analogue of (7.24) for free boundary conditions is actually false. Technically, the way the BC decoupling inequality is used to establish (7.24) is to show that the probabilities of the finite-volume events defining $\{x \leftrightarrow y\}$ are monotone in the volume. A second type of application of the BC inequality is less technical, and will be discussed in some detail in Subsection 7.7.

7.5 Comparison Principles

The comparison principles are another set of non-standard correlation inequalities for the random cluster model. They allow us to compare Potts models at different values of p and q. Given these principles, if we can prove the existence of a phase transition at one value of q, we will automatically have a similar result for other values of q. However, our estimates on the temperature at which the transition occurs will deteriorate.

DEFINITION: Let $(\Omega, \mathcal{F}, \mu)$ and $(\Omega, \mathcal{F}, \nu)$ be probability spaces, and suppose there exists a partial order on Ω. We say that μ *FKG dominates* ν and write $\mu \underset{\rm FKG}{\geq} \nu$ if

$$\mu(f) \geq \nu(f)$$

for all increasing functions $f : \Omega \to \mathbb{R}$.

In the theorem below, we will use $\mu_{p,q}$ to denote a random cluster measure with parameter values q and $p = \{p_b\}$. The inequality $p_1 \leq p_2$ will mean $p_b^1 \leq p_b^2$ for every bond $b \in \mathbb{K}_d$.

THEOREM 7.4. ([ACCN2]) *Suppose μ_{p_1,q_1} and μ_{p_2,q_2} are both free or both wired (infinite-volume) random cluster measures. Let $q_1 \geq q_2 \geq 1$.*
(i) If $p_1 \leq p_2$, then

$$\mu_{p_1,q_1} \underset{\rm FKG}{\leq} \mu_{p_2,q_2} ,$$

(ii) while if $\frac{p_1}{q_1(1-p_1)} \geq \frac{p_2}{q_2(1-p_2)}$, then

$$\mu_{p_1,q_1} \underset{\rm FKG}{\geq} \mu_{p_2,q_2} .$$

PROOF. First note that it suffices to establish the dominance relations for finite-volume measures, since these relations are inherited by the corresponding infinite-volume measures.

We will use the fact that the measure μ is FKG if for all increasing events A, $B \in \mathcal{F}$,

$$\mu(A \cap B) \geq \mu(A)\mu(B).$$

Equivalently, μ is FKG if for every nonnegative, increasing function f with finite expectation $\mu(f) < \infty$,

$$\frac{\mu(\cdot f)}{\mu(f)} \underset{\text{FKG}}{\geq} \mu(\cdot).$$

Note that, in (i), $q_1 \geq q_2 \geq 1$ so that μ_{p_1,q_1} is FKG, while, in (ii), $q_2 \geq 1$ so that μ_{p_2,q_2} is FKG. In order to simplify notation, let $\mu_1 = \mu_{p_1,q_1}$ and $\mu_2 = \mu_{p_2,q_2}$.

Now for either μ,

$$\begin{aligned}\mu &\propto q^{\#(\omega)} \prod_b p_b^{\omega_b}(1-p_b)^{1-\omega_b} \\ &\propto q^{\#(\omega)} \prod_b \left(\frac{p_b}{1-p_b}\right)^{\omega_b}.\end{aligned}$$

(Note that this product makes sense since we are assuming μ is a finite-volume measure.) Thus μ_1 can be written as

$$\mu_1(\cdot) = \frac{\mu_2(\cdot f)}{\mu_2(f)}$$

with

$$\begin{aligned}f(\omega) &= \left(\frac{q_1}{q_2}\right)^{\#(\omega)} \prod_b \left(\frac{p_b^1/(1-p_b^1)}{p_b^2/(1-p_b^2)}\right)^{\omega_b} \\ &= \left(\frac{q_1}{q_2}\right)^{\#(\omega)+\sum_b \omega_b} \prod_b \left(\frac{p_b^1/(q_1(1-p_b^1))}{p_b^2/(q_2(1-p_b^2))}\right)^{\omega_b}.\end{aligned}$$

Observe that

(a) $\#(\omega)$ is decreasing in ω since occupying more bonds might connect clusters, while

(b) $\#(\omega) + \sum_b \omega_b$ is increasing ω since occupying a single bond decreases $\#(\omega)$ by at most 1 but increases $\sum_b \omega_b$ by 1.

Hence, in case (i), f is decreasing, while in case (ii), f is increasing, which proves the theorem. $\square$

Now let us briefly discuss the applications of the comparison principles to establish the existence of phase transitions in certain Potts models. These applications are a consequence of the following corollary.

Corollary 7.5. ([ACCN2]) *Suppose $\{J_{x,y}\}$ is a fixed set of couplings. Let $\beta_t(q)$ denote the inverse transition temperature defined in (7.6) for the random cluster model with parameter q. Then*

$$\beta_t(q') \geq \beta_t(q) \geq \frac{q}{q'}\beta_t(q')$$

for all $q' \geq q \geq 1$.

Proof. Using the random cluster expression (7.12) for the magentization, the corollary is a straightforward application of the two comparison principles. $\square$

This implies that whenever we know the existence of a phase transition at a model with one value of $q \geq 1$ (e.g. percolation), we automatically know it at all others, although our estimate of the transition temperature will deteriorate by the ratio of the q's. In particular, in [ACCN2] it was pointed out that this corollary, combined with previous results, immediately leads to a proof of a phase transition in the following systems:

Long-range One-dimensional Models.

Consider a one-dimensional model with long-range ferromagnetic couplings

$$J_{x,y} \approx c/|x-y|^s,$$

$c \geq 0$. The proof of existence of a phase transition and the nature of the phase transition in Ising models with couplings of this form, particularly the model with $s = 2$, was the subject of intense investigation for over 15 years. Dobrushin ([Do1], [Do2]), Ruelle [Rue], and Dyson ([Dy1], [Dy2]) established that the critical power for a phase transition is $s = 2$: [Do1], [Do2] and [Rue] showed that there is no long-range order if $J_{0,x} \approx |x|^{-s}$ with $s > 2$, while [Dy1], [Dy2] showed that for any $s < 2$, long-range order will occur for sufficiently large β. That a phase transition actually occurs in the delicate $s = 2$ case was finally established by Fröhlich and Spencer [FrS], inspired by the renormalization group analysis of Anderson, Yuval, and Hamann [AYH]. Meanwhile, Newman and Schulman [NS2] showed the existence of percolation in $1/x^2$ systems. We now realize that the [FrS] and [NS2] results can be obtained from each other using Corollary 7.5. Moreover, we can also automatically establish the existence of a transition for any $q \geq 1$ from either of these results.

Next, motivated by Thouless' prediction [Th] for the $1/x^2$ Ising model, Aizenmann and Newman [AN2] established the nature of the phase transition in $1/x^2$ percolation model. Namely, they showed that the percolation probability is discontinuous at the transition point (i.e., $P_\infty(p_c) > 0$). This jump is known as the Thouless effect. It is remarkable in that the system behaves in most respects like a system with a second-order phase transition (e.g., the correlation length diverges at the transtion point), but nevertheless the order parameter is discontinuous. Finally, using the techniques discussed here—although unfortunately not simply the corollary—Aizenmann, Chayes, Chayes, and Newman [ACCN2] proved the Thouless effect for all the $q \geq 1$ state models with $J_{0,x} \approx c/x^2$. In fact, many of the techniques discussed here were developed to solve that problem.

Dilute and Random Ferromagnets.

An important and surprisingly easy application of the corollary is to study the phase transitions in dilute and random Ising and Potts ferromagnets. As described earlier, these are models in which the coupling $J_{x,y}$ are i.i.d. variables distributed according to some distribution $\rho(J)$ supported entirely on non-negative J. They are called *dilute ferromagnets* if ρ has an atom at $J = 0$; otherwise, they are just said to be *random ferromagnets*. By the corollary, if we can prove the existence of a phase transition in such a system with one value of $q \geq 1$, then we have it for all others. Obviously, the easiest case is $q = 1$. For example, if we look at simple bond dilution (i.e., $\rho(J)$ has two atoms: one at $J = 0$ and another at $J = J_0 > 0$), the effect of the dilution is just a shift in the effective bond density, so that we again have a pure percoaltion model. More complicated distributions are also quite tractable. Not only can we use the corollary to establish the existence of a phase transition in all the corresponding Ising and Potts models. We can also use these methods to study the so-called crossover, when the dilution density is close to the natural threshold, see [ACCN1].

Percolation on Wedges.

Grimmett [Grim1] studied nearest-neighbor bond percolation on slices of $\mathbb{Z}^2$ of the form

$$S_f = \{x \in \mathbb{Z}^2 : x_1 > 0, |x_2| \leq f(x_1)\}$$

where f is some strictly positive function. He showed that if

$$f(z) \approx a \log(z),$$

then percolation occurs on S_f, but $p_c(a) \to 1$ as $a \to 0$. These results were extended to higher dimensions by Hammersley and Whittington [HW] and to Ising systems by Chayes and Chayes [CC2]. [CC2] also showed that $\pi_c \neq p_c$ for percolation on these logarithmic wedges. Using the corollary and the aforementioned results, it follows that if we consider ferromagnetic nearest-neighbor $q \geq 1$ state Potts models on

$$S_f = \{x \in \mathbb{Z}^d : x_1 \geq 0, \max_{i=1,\dots d} |x_i| \leq f(x_1)\}$$

with

$$f(z) \approx (a \log z)^{\frac{1}{d-1}},$$

then spontaneous magnetization occurs on S_f, but $\beta_c(a) \to \infty$ as $a \to 0$. On the other hand, provided that $\lim_{z\to\infty} f(z) = \infty$, the analogue of π_c in the system (i.e., the point at which the correlation length diverges) coincides with that of the full lattice $\mathbb{Z}^d$ at the same temperature. Thus there is an intermediate phase in logarithmically growing wedges.

7.6 The DLR Equation and States of the Random Cluster Model

In this subsection, we discuss states of the random cluster model. First, we follow almost verbatim the treatment of Borgs and Chayes [BC] and introduce the notion of DLR (Dobrushin-Lanford-Ruelle) states for the random cluster model. We then review the proof of Grimmett [Gr3] and Pfister and Vande Velde [PV] that the free and wired random cluster measures are DLR states. Finally, we return to the [BC] treatment, showing that the DLR condition implies ergodicity of the free and wired states, and drawing some other conclusions about the structure of states in the random cluster model.

However, before discussing DLR states of the random cluster model, let us pause to discuss the general notion of DLR states. Given an arbitrary spin model, one would like to find the smallest set of states $\mathcal{G}$ having the following properties: (1) $\mathcal{G}$

contains all limits of finite-volume Gibbs states that can be obtained with various boundary conditions; (2) $\mathcal{G}$ is convex, i.e. closed under "statistical mixtures"; and (3) $\mathcal{G}$ is closed under weak limits. Dobrushin ([Do1], [Do2]) and, independently, Lanford and Ruelle [LR], proposed conditions which can be shown to lead to such a set of states. These states are now known either as DLR states or equilibrium states or (infinite-volume) Gibb states; the three terms are usually used interchangeably. Basically, Dobrushin, Lanford and Ruelle showed that equilibrium states are precisely those which have the correct finite-volume conditional probabilities, where the conditioning is the result of the boundary condition induced on the finite volume by the infinite-volume state. (That is, we draw configurations according to the infinite-volume measure and use these configurations to specify boundary conditions on a finite volume.) Thus, in order to define DLR states for the random cluster model, we must see how measures in a larger volume induce boundary conditions on a smaller volume, and we must guarantee that the measures so induced are consistent with each other. This is precisely what is done in [BC], and is the treatment we follow here.

The DLR condition is the infinite-volume analogue of the consistency condition described above. One usually shows existence of infinite-volume measures satisfying the DLR condition by invoking the general theory of Gibbs states (see e.g. [Pr] and [Ge]). However, the general theory requires that the finite-volume expectations used to construct the DLR equations are *quasilocal* functions of the boundary conditions. Below we will define quasilocality and show that it fails to hold here due to the nonlocality of the random cluster weights. Thus the DLR equation has to be established explicitly. This is done for the free state by Borgs and Chayes [BC], and for both the free and wired states by Grimmett [Grim3] and Pfister and Vande Velde [PV]. The crux of the Grimmett and Pfister-Vande Velde proof is to use uniqueness of the infinite cluster to show that although the specification is not pointwise quasilocal, it is in fact almost surely quasilocal.

Here we will review uniqueness and almost sure quasilocality. We then omit the explicit proof of the DLR condition, since given a.e. quasilocality, the remainder of the proof is rather standard. Finally, we will return to the treatment of [BC], showing how the DLR condition implies ergodicity of the free and wired states, and drawing some other conclusions about the structure of states in the random cluster model.

For simplicity, in this subsection we will restrict attention to the nearest-neighbor random cluster model on the hypercubic lattice with uniform bond density, i.e. here $\mathbb{L}_d = \mathbb{Z}^d$, we consider $\mathbb{B}_d$ instead of $\mathbb{K}_d$, and $p_b = p$ for all b. We let $B(\Lambda)$ denote the nearest-neighbor bond graph on Λ and $B^+(\Lambda)$ denote $B(\Lambda)$ plus the bonds which connect Λ to $\partial\Lambda_1$:

$$B^+(\Lambda) = \{\langle x, y\rangle : |x-y| = 1, x \in \Lambda, \text{ and } y \in \Lambda \cup \partial\Lambda_1\}\,.$$

We start by defining finite-volume measures with general unconstrained boundary conditions—conditions which permit any component to be connected to any other component (see the second remark at the end of this subsection). The set of states generated by all such boundary conditions is quite natural in the random cluster model. A larger class including constrained states will be discussed briefly at the end of this subsection. Each measure is defined on an arbitrary finite set of bonds $B \subset \mathbb{B}_d$ with boundary

$$\partial B = \{x \in \mathbb{Z}^d \mid \exists y, z \in \mathbb{Z}^d \text{ with } \langle x,y\rangle \in B\,, \langle x,z\rangle \in B^c\}\,.$$

[BC] specify the boundary condition by introducing a *wiring diagram*, W, which is a disjoint partition of ∂B into $n_W = 1, \cdots, |\partial B|$ components:

$$W = \{W_1, \cdots, W_{n_W}\} \quad \text{with} \quad \partial B = \bigcup_{i=1}^{n_W} W_i\ , \quad W_i \cap W_j = \emptyset \text{ if } i \neq j\,.$$

They denote by $\mathcal{W}(\partial B)$ the set of all such wiring diagrams—i.e. the set of all disjoint partitions of ∂B. Each component, W_i, of the wiring diagram W is considered to be preconnected or wired, so that all bonds $b \in B$ connected to points of W_i are regarded as being connected to each other. The number of components $\#(\omega)$ is then computed as usual. The random cluster weight

$$G_{W,B}(\omega) = \frac{1}{Z_W(B)}(1-p)^{|\omega^c|}p^{|\omega|}q^{\#(\omega)} \tag{7.25}$$

defines the finite-volume measure $\mu_{W,B}(\cdot)$. Denoting by $W_{\rm free}$ the partition with $n_W = |\partial B|$ components and by $W_{\rm wir}$ the partition with only a single component, we see that $\mu_{{\rm free},\Lambda}(\cdot) = \mu_{W_{\rm free},B(\Lambda)}(\cdot)$ and $\mu_{{\rm wir},\Lambda}(\cdot) = \mu_{W_{\rm wir},B^+(\Lambda)}(\cdot)$, so that the free and (fully) wired measures are just special cases of $\mu_{W,B}(\cdot)$. Note that among the measures $\mu_{W,B}(\cdot)$ are some that cannot be obtained as transforms of any finite-volume states in the spin system, namely those in which W has more than q components W_i with $|W_i| \geq 2$.

There is a natural partial order on the set $\mathcal{W}(\partial B)$. If $W, W' \in \mathcal{W}(\partial B)$, we say that W' is *coarser* than W, denoted by $W' \succeq W$, if for each $W_i' \in W'$ there exist $W_{i_1}, W_{i_2}, \cdots, W_{i_m} \in W$ such that $W_i' = \cup_{j=1}^m W_{i_j}$. Notice that $W_{\rm free}$ is the least coarse and $W_{\rm wir}$ is the most coarse of all wiring diagrams. Moreover if $W' \succeq W$, then $\mu_{W',B}$ dominates $\mu_{W,B}$ in the sense of FKG.

One of the fundamental ideas in the theory of DLR states is that configurations on larger sets naturally induce boundary conditions on smaller sets. Here, each configuration $\omega \in \Omega$ induces a wiring diagram on each finite set $B \subset \mathbb{B}_d$. The *induced wiring diagram* $W(B,\omega)$ is a partition into components of ∂B, each of which is connected using occupied bonds in ω_{B^c}. Thus each $\omega \in \Omega$ gives rise to a sequence of induced finite-volume measures $\mu_{W(B,\omega),B}$ for any increasing sequence of sets $B \subset \mathbb{B}_d$. Henceforth we will extend the induced finite-volume measure $\mu_{W(B,\omega),B}$ to a measure on the full space $(\Omega, \mathcal{F})$ by declaring all bonds in B^c agree with the configuration specified by ω. Using the form (7.25) of the weights $G_{W,B}$ and the [BC] definition of induced wiring diagrams, it is straightforward to check that the (extended) induced finite-volume measures obey the *consistency condition*

$$\mu_{W(B,\omega),B}(A) = \int \mu_{W(B,\omega),B}(d\tilde{\omega})\ \ \mu_{W(\tilde{B},\tilde{\omega}),\tilde{B}}(A) \tag{7.26}$$

for all local events $A \in \mathcal{F}$ (i.e. events which depend on only a finite number of bonds), any finite set B and all $\tilde{B} \subset B$. This consistency condition says that the expectation of A on the larger set is obtained by averaging its conditional expectation over smaller sets, where the conditioning is precisely the boundary condition induced by the configuration on the larger set. In other words, the configuration on the smaller set is in equilibrium with its exterior.

Let $\omega \in \Omega$. For each finite B, we may extend the random cluster measure $\mu_{W(B,\omega),B}(\cdot)$ to a measure $\pi_B(\cdot,\omega)$ on $\mathcal{F}$ by requiring
(i) $\forall A \in \mathcal{F}_B$, $\pi_B(A,\omega) = \mu_{W(B,\omega),B}(A)$,
(ii) $\forall A \in \mathcal{F}_{B^c}$, $\pi_B(A,\omega) = \mathbf{1}_A(\omega)$, where $\mathbf{1}_A$ is the indicator function of A, and
(iii) $\pi_B(\cdot,\omega)$ is a product measure on $\mathcal{F}_B \times \mathcal{F}_{B^c}$.

For each finite B, we therefore have a function

$$\pi_B : (\mathcal{F}, \Omega) \to \mathbb{R}\,.$$

Consider the family

$$\gamma = \{\pi_B | B \subset \mathbb{B}_d,\ |B| < \infty\}\,. \tag{7.27}$$

By the consistency condition (7.26), the family γ is what is called a *specification* in the sense of [Pr]. (By definition, a specification is simply a family of probability kernels obeying a measurability condition, an indicator condition of the form (ii) and a consistency condition of the form (7.26). Here the measurability condition amounts to the requirement that for each $A \in \mathcal{F}$, $\pi_B(A,\omega)$ depends only on the part of the configuration ω in B^c.)

A DLR equation ([Do1], [Do2], [LR]) is just an infinite-volume analogue of a consistency condition like (7.26). Thus we introduce the (unconstrained) DLR equation for an infinite-volume random cluster state μ:

$$\mu(A) = \int \mu(d\omega)\ \ \mu_{W(B,\omega),B}(A) \tag{7.28}$$

where $A \in \mathcal{F}$ is any local event and $B \subset \mathbb{B}_d$ is any finite set. As usual, the DLR equation (7.28)—if it holds—allows us to write the infinite-volume expectation of A as an average over finite-volume expectations. It is closed in the sense that the average is computed with respect to the given measure μ.

Let us denote the set of states obeying (7.28) by $\mathcal{G} = \mathcal{G}(\gamma)$, where as above γ denotes the specification. States $\mu \in \mathcal{G}$ will be called *DLR states* or *Gibbs states* or *equilibrium states*. *A priori* it is not clear whether $\mathcal{G}$ is nonempty, i.e. whether there exists any μ satisfying (7.28). One might try to construct such a μ as a subsequential limit of finite-volume measures $\mu_{W,B}$—which clearly exists by compactness—but the question of whether such a limit obeys (7.28) involves a delicate interchange of limits. The theory of Gibbs states ([Pr], [Ge]) provides general conditions under which (7.28) is satisfied, one of which is quasilocality of the specification.

A function f is *quasilocal* if it can be approximated in the supremum norm by local functions, a property which is equivalent ([Ge], Remark 2.21) to the statement

$$\sup_{\omega,\eta\,:\,\omega_B=\eta_B} |\ f(\omega) - f(\eta)\ | \to 0 \quad \text{as } B\ \to \mathbb{B}_d\ .$$

A specification $\{\pi_B\}$ is quasilocal if the functions $\pi_B(A,\cdot)$ are quasilocal for all finite $B \subset \mathbb{B}_d$ and all local events $A \in \mathcal{F}$.

Unfortunately, due to nonlocality of the weights $G_{W,B}$, our specification is not quasilocal. For example, the probability of the simple event $\{\omega_b = 1\}$, conditioned on the bonds in $\mathbb{B}_d \setminus \{b\}$, changes discontinuously depending on whether or not the endpoints of b are connected by a path (of any length) in $\mathbb{B}_d \setminus \{b\}$ (see (7.18) or next paragraph). The general theory of Gibbs states therefore can not be applied here. [BC] circumvented this difficulty by explicitly showing that the measure μ_{free} is a Gibbs measure.

Here we follow instead the tack of Grimmett [Grim3] and Pfister and Vande Velde [PV] and show that uniqueness of the infinite cluster implies almost sure quasilocality of the specification. In order to prove uniqueness, we will use the Burton and Keane theorem. To this end, we first note that both the free and wired measures, μ_{free} and μ_{wir}, are stationary. Thus it remains to establish finite energy. Now, it is easy to see that finite energy is equivalent to the statement: for each bond b, the conditional probability of the event that b is occupied, given the configuration on all the other bonds, is nontrivial:

$$0 < \mu(\omega_b = 1 \,|\, \omega_{\tilde{b}}, \tilde{b} \neq b) < 1 .$$

For the free and wired measures, it was observed already that this probability can be explicitly calculated:

$$\mu(\omega_b = 1 \,|\, \omega_{\tilde{b}}, \tilde{b} \neq b) = \begin{cases} p & \text{if the endpoints of } b \text{ are connected, regardless of } \omega(b) \\ \frac{p}{p+q(1-p)} & \text{otherwise,} \end{cases}$$

where $\mu = \mu_{\text{free}}$ or μ_{wir}. Thus for all $q \geq 1$ and all $p \neq 0, 1$, the random cluster measures μ_{free} and μ_{wir} have finite energy. Note that this is not true in all random cluster measures: Boundary conditions can impose constraints which exclude certain configurations.

Let $\mathcal{S}_{p,q}$ the closed convex hull of the set of all stationary infinite-volume random cluster measures at parameters p and q which can be obtained as the weak limits of induced finite-volume measures $\mu_{W(B,\omega),B}$ for some configuration ω. Then each $\mu \in \mathcal{S}_{p,q}$ is stationary and has finite energy, and thus obeys the Burton and Keane [BuK] theorem:

PROPOSITION 7.6. *For any $q \geq 1$ and any $p \in (0,1)$, any random cluster state $\mu \in \mathcal{S}_{p,q}$ has at most one infinite cluster with probability one. In particular, the free and wired states have at most one infinite cluster with probability one.*

Since the basic Burton and Keane theorem requires only stationarity, it applies also to non-extremal states, and therefore allows the possibility of a convex combination of states with zero and one infinite cluster. If, in addition, the measures are ergodic, then at any given value of p, there is *either* zero *or* one infinite cluster with probability one. It will turn out that this is the case for both the free and wired measures, although in order to prove it, we will have to use uniqueness to establish a.s. quasilocality and thus satisfaction of the DLR condition (which together with the FKG ordering will then imply extremality and ergodicity). Of course, ergodicity does not exclude the possibility that, for a fixed value of p, the wired state has an infinite cluster and the free state does not—indeed, for q large enough, this is exactly what happens at the transition point.

THEOREM 7.7. ([Grim3], [PV]) *Let $q \geq 1$ and $p \in (0,1)$. Let $\mu \in \mathcal{S}_{p,q}$. Then the specification γ defined in (7.27) is Ω-a.s. quasilocal.*

PROOF. We follow the proof of [Grim3]. Let $g(\omega) = \mu_{W(B,\omega),B}(A)$ and define the corresponding "discontinuity set" by

$$D = \bigcap_{\Delta} \left\{ \omega : \sup_{\eta : \eta_\Delta = \omega_\Delta} |g(\eta) - g(\omega)| > 0 \right\}$$

where the intersection is over all boxes Δ containing B. Now let $D_{\Delta,B}$ be the event that two points $x, y \in \partial B$ are each connected to $\partial\Delta$ by occupied paths of ω in the annulus $\Delta \backslash B$, but these two paths are not connected to each other. It is easy to see that $|g(\eta) - g(\omega)| > 0$ only if such an event $D_{\Delta,B}$ occurs. Thus

$$D \subset \bigcap_{\Delta} D_{\Delta,B}\,,$$

so that

$$\mu(D) \leq \mu\left(\bigcap_{\Delta} D_{\Delta,B}\right).$$

But the right-hand side is zero since $\bigcap_{\Delta} D_{\Delta,B}$ can only occur if there are at least two infinite clusters. □

It is now rather standard to show that μ_{free} and μ_{wir} satisfy the DLR equation. We have

PROPOSITION 7.8. ([Grim3]) *Let $q \geq 1$ and $p \in [0,1]$. Then $\mu_{\text{free}} \in \mathcal{G}$ and $\mu_{\text{wir}} \in \mathcal{G}$.*

Moreover, this easily implies that these measures are ergodic:

THEOREM 7.9. ([BC]) *Let H be any nontrivial subgroup of the translation group and let $\mathcal{G}_o \subset \mathcal{G}$ be the set of all H-invariant DLR states. Then for all $q \geq 1$, μ_{free} and μ_{wir} are extremal in $\mathcal{G}_o$ and hence are H-ergodic.*

PROOF. As noted earlier, W_{free} is the least coarse and W_{wir} is the most coarse of all wiring diagrams, so that

$$\mu_{W_{\text{free}},B} \underset{\text{FKG}}{\leq} \mu_{W,B} \underset{\text{FKG}}{\leq} \mu_{W_{\text{wir}},B} \qquad \text{for all } W \in \mathcal{W}(\partial B)\,,$$

and thus by convergence of the measure (7.11)

$$\mu_{\text{free}} \underset{\text{FKG}}{\leq} \mu \underset{\text{FKG}}{\leq} \mu_{\text{wir}} \qquad \text{for all } \mu \in \mathcal{G}\,. \tag{7.29}$$

Given that $\mu_{\text{free}}, \mu_{\text{wir}} \in \mathcal{G}$ by Proposition 7.8, it follows immediately from (7.29) that μ_{free} and μ_{wir} are extremal in $\mathcal{G}$ and hence also in $\mathcal{G}_o$ (since the measures are of course H-invariant). Ergodicity then follows from the fact that all extremal measures in $\mathcal{G}_o$ are H-ergodic ([Pr], Theorem 4.1). $\square$

REMARKS. (i) *The Size of* $\mathcal{G}$: By Proposition 7.8, μ_{free} and μ_{wir} are in $\mathcal{G}$ so that $|\mathcal{G}| \geq 1$ for all $q \geq 1$ and all inverse temperatures β. According to a result of [ACCN2] (Theorem A.2), whenever $M(\beta) = 0$ (i.e. $\beta \leq \beta_t$ for systems with second-order transitions and $\beta < \beta_t$ for those with first-order transitions) $\mu_{\text{free}} = \mu_{\text{wir}}$, so that by (7.29) and Proposition 7.8, $|\mathcal{G}| = 1$. It is expected that $|\mathcal{G}| = 1$ also for $\beta > \beta_t$, but there are only incomplete results for $d = 2$: The two-dimensional dual of the [ACCN2] result says $M(\beta^*) = 0$ implies $|\mathcal{G}| = 1$, i.e. there is one state for $\beta > \beta_t^*$, which presumably coincides with β_t. (Recall that we expect $\beta_t = \beta_s$ in $d = 2$; see equations (7.3) and (7.6).)

However, one expects more states at the transition point in systems with first-order transitions. For q large enough and $d = 2$, convergent expansions ([KoS], [LMR]) can be used to show that there are $q + 1$ distinct translation-invariant spin states (which transform into two distinct translation-invariant random cluster states—the free and the wired). There are presumably no non-translation-invariant states. Thus we expect $|\mathcal{G}| = 2$ for $\beta = \beta_t$ and q large enough in $d = 2$. In three dimensions, convergent expansions [MMRS] can be used to show that for q large enough, in addition to the translation-invariant states discussed above, there are infinitely many non-translation-invariant "Dobrushin-type" states corresponding here to states constructed from wiring diagrams which coincide with W_{wir} above a certain hyperplane and with W_{free} below that plane. We expect that these expansions can also be used to show that these non-translation-invariant states satisfy the DLR equation (7.28), so that at $\beta = \beta_t$, $|\mathcal{G}| = \infty$ for q large enough in $d \geq 3$.

(ii) *States with Constraints:* In the remark above, we mentioned "Dobrushin-type" states which we expect to be in $\mathcal{G}$; these states were constructed from a combination of wired and free boundary conditions. There are, however, many Dobrushin-type states in the spin system whose transforms are not in $\mathcal{G}$: namely, mixed states in which various components of the boundary have different values of the spin. In the random cluster model, these correspond to states with constraints—certain components cannot be connected to other components. Therefore, in order to formulate DLR equations for these states, one has to supplement our wiring diagrams with some notion of constraints. While this is possible for individual finite-volume states, it is not clear how constraints should be induced by a given configuration $\omega \in \Omega$, nor whether the resulting measures would obey even finite-volume consistency conditions.

7.7 Length Scales in the Potts Models

The final subsection of these notes concerns an application of the FK representation and the methods reviewed here, particularly the decoupling inequalities, to the question of length scales in the two-dimensional Potts models. As we already discussed in Subsection 7.1, for q sufficiently large (presumably for $q > 4$ in $d = 2$), the q-state Potts model has a first-order transition, so that in particular the correlation length is finite at the transition temperature: $\xi(\beta_t) < \infty$. The results of this subsection were motivated by a discrepancy which arose in the determination of $\xi(\beta_t)$.

In 1993, the correlation length at the self-dual point β_s, cf. (7.3), was explicitly calculated using a mapping of the critical Potts model into the exactly solvable six-vertex model ([BW], see also [KSZ], [Kl]). However, this explicit value was roughly twice the value measured in previous numerical experiments ([PL], [GI]). In an attempt to resolve the apparent inconsistency, Borgs and Janke [BJ] suggested that the exact calculations might be relevant for the correlation length $\xi_-(\beta)$ coming from the low-β, or disordered, phase, while the numerical work might be measuring the correlation length $\xi_+(\beta)$ coming from the high-β, or ordered, phase. Furthermore, they suggested the relation

$$\xi_+(\beta_s) = \frac{1}{2}\xi_-(\beta_s). \tag{7.30}$$

A continuous transition analogue of this relation was already known for two-dimensional bond percolation, where

$$\xi(p) = \frac{1}{2}\xi(1-p) \tag{7.31}$$

for all $p > p_c$ was rigorously established by [CCGKS], and for the two-dimensional Ising magnet, where

$$\xi(\beta) = \frac{1}{2}\xi(\beta^*) \tag{7.32}$$

for all $\beta > \beta_s$ was established via exact solution by [MW]. However, it was initially quite unexpected to have a relation of this form at a discontinuous transition.

Let us contrast the situations. At a second-order (continuous) transition, the correlation length diverges coming both from above and below threshold.

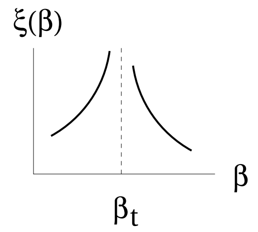


FIGURE 7.5. *The expected behavior of ξ in a system with a second-order transition.*

As we have learned, one typically assumes power laws:

$$\begin{aligned} \xi(\beta) &\approx \xi_+|\beta-\beta_t|^{-\nu_+} \quad \text{as} \quad \beta \searrow \beta_+ \\ \xi(\beta) &\approx \xi_-|\beta-\beta_t|^{-\nu_-} \quad \text{as} \quad \beta \nearrow \beta_-. \end{aligned}$$

Not only is it expected that

$$\nu_+ = \nu_-,$$

(which is Widom scaling in two dimensions), but also that the amplitude ratio $\xi_+(\beta_t)/\xi_-(\beta_t)$ is universal, i.e. $\xi_+(\beta_t)/\xi_-(\beta_t)$ depends only on the universality class of the model. This is expected to hold in all dimensions. The two-dimensional relations (7.31) and (7.32) of course guarantee that $\nu_+ = \nu_-$.

At a first-order (discontinuous) transition, the correlation length coming from either direction is finite. Clearly, there is no reason for these lengths to be equal. In fact, since there is no universality at first-order transitions, it was expected that $\xi_+(\beta_t)$ and $\xi_-(\beta_t)$ should be entirely unrelated.

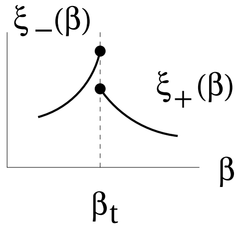


FIGURE 7.6. *The expected behavior of* ξ_+ *and* ξ_- *in a system with a first-order transition.*

The insight of [BC] was that maybe the percolation and Ising relations, (7.31) and (7.32), are a consequence not of universality, but of two-dimensional duality, so that such a relation could also hold for models with two-dimensional discontinuous transitions. This turned out to be the case. The proof required extensive use of the FK representation, including the development of the decoupling inequalities to extend the percolation proof of relation (7.31) to the random cluster model, which had neither independence nor a BK inequality.

The first step in the proof of the relation (7.30) is to determine exactly what we mean by the correlation length. This part of the proof holds in all dimensions, not just in $d = 2$. Clearly, the correlation length is the decay rate of some two-point function. But which two-point function should we choose? For each set of boundary conditions φ, [BC] considered the full covariance matrix

$$G^{mn}_{\varphi}(x-y) = \langle q\delta(\sigma_x, m); q\delta(\sigma_y, n)\rangle_{\varphi}$$

where $\langle A; B\rangle_{\varphi} = \langle AB\rangle_{\varphi} - \langle A\rangle_{\varphi}\langle B\rangle_{\varphi}$ is the truncated expectation of the functions A and B. The nontrivial eigenvalues of this $q \times q$ matrix are the invariant two-point functions, and their decay rates are therefore the correlation lengths.

In the disordered phase (i.e. for $\beta < \beta_t$), it is appropriate to consider the covariance matrix with free boundary conditions, $G^{mn}_{\text{free}}(x-y)$. For $q \geq 1$, [BC] find that this is proportional to the standard two-point function, which in turn is equal to the connectivity function in the random cluster representation:

$$\begin{aligned} G^{mn}_{\mathrm{free}}(x-y) &= (q\delta(m,n)-1)\langle\frac{1}{q-1}(q\delta(\sigma_x,\sigma_y)-1)\rangle_{\mathrm{free}} \\ &= (q\delta(m,n)-1)\tau_{\mathrm{free}}(x-y)\,, \end{aligned} \tag{7.33}$$

see equation (7.13). Here the connectivity, $\tau_{\mathrm{free}}(x-y)=\tau_{\mathrm{free}}(x,y)$, is the probability with respect to the free boundary condition random cluster measure that x and y lie in the same component. It is easy to establish relation (7.33) in a finite volume. Taking the infinite-volume limit requires monotonicity, which is guaranteed by the FKG inequality. We note that, in the disordered phase, the covariance matrix contains no more information than the standard two-point function, or equivalently, the connectivity function. The matrix (7.33) can easily be diagonalized, yielding a simple eigenvalue zero and a $(q-1)$-fold degenerate eigenvalue

$$G_{\mathrm{free}}(x-y) = q\tau_{\mathrm{free}}(x-y). \tag{7.34}$$

The problem is more subtle in the ordered phase, where it is appropriate to consider the matrix $G^{mn}_c(x-y)$ with fixed constant boundary conditions, $c\in S=\{1,\cdots,q\}$. Defining the finite-cluster connectivity, $\tau^{\mathrm{fin}}_{\mathrm{wir}}(x-y)=\tau^{\mathrm{fin}}_{\mathrm{wir}}(x,y)$, to be the probability, in the wired random cluster measure, that x and y lie in the same *finite* component, and the infinite-cluster covariance, $C_{\mathrm{wir}}(x-y)$, to be the covariance, again in the wired measure, of the events that x and y lie in the infinite component, [BC] prove that for $q\geq 1$ the matrix elements $G^{mn}_c(x-y)$ are linear combinations of $\tau^{\mathrm{fin}}_{\mathrm{wir}}(x-y)$ and $C_{\mathrm{wir}}(x-y)$, namely

$$G^{mn}_c(x-y) = (q\delta(m,n)-1)\,\tau^{\mathrm{fin}}_{\mathrm{wir}}(x-y) + (q\delta(m,c)-1)(q\delta(n,c)-1)C_{\mathrm{wir}}(x-y)\,. \tag{7.35}$$

We remark that while the finite-volume analogue of (7.35) is a straightforward consequence of the FK representation, the proof of the infinite-volume limit involves some subtleties related to how the infinite cluster emerges from large finite clusters in the wired problem. This time, the required monotonicity is a consequence of the BC decoupling inequalities (see the discussion at the end of Subsection 7.5, in particular equation (7.24)).

Percolation analogues of $\tau^{\mathrm{fin}}_{\mathrm{wir}}(x-y)$ and $C_{\mathrm{wir}}(x-y)$—in the absence of boundary conditions—arose previously in the percolation proof of [CCGKS], where they appeared as a natural decomposition of the truncated percolation connectivity in the ordered phase. There, however, they did not have independent significance, appearing only as a sum. The question naturally arises whether they have independent significance here. Obviously, this is not the case for $q=2$, for which (7.35) can be rewritten as

$$G^{mn}_c(x-y) = (2\delta(m,n)-1)\,(\tau^{\mathrm{fin}}_{\mathrm{wir}}(x-y)+C_{\mathrm{wir}}(x-y))\,,$$

involving again only the sum $\tau^{\mathrm{fin}}_{\mathrm{wir}}(x-y)+C_{\mathrm{wir}}(x-y)$.

For $q \geq 3$, however, the fixed boundary condition covariance matrix $G_c^{mn}(x-y)$ has a richer structure. [BC] prove that it has a simple eigenvalue zero and a nontrivial simple eigenvalue

$$G_{\mathrm{wir}}^{(1)}(x-y) = q\, \tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y) + q(q-1)C_{\mathrm{wir}}(x-y)\,, \tag{7.36}$$

both corresponding to the trivial representation of the unbroken subgroup S_{q-1} of permutations of $S \setminus \{c\}$, as well as one $(q-2)$-fold degenerate eigenvalue

$$G_{\mathrm{wir}}^{(2)}(x-y) = q\, \tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)\,, \tag{7.37}$$

corresponding to the remaining orthogonal subspace.[1] Thus we see that for $q \geq 3$, the finite-cluster cluster connectivity, $\tau_{\mathrm{wir}}^{\mathrm{fin}}(x-y)$, has independent algebraic significance as an eigenvalue of the covariance matrix, and hence also physical significance in terms of the associated one-particle spectrum. As for the infinite cluster covariance $C_{\mathrm{wir}}(x-y)$, it can be shown ([BC], Theorem 4.3) that its *decay rate* is equal to the *decay rate* of the eigenvalue $G_{\mathrm{wir}}^{(1)}$ whenever the magnetization is positive. Thus although $C_{\mathrm{wir}}(x-y)$ does not have independent algebraic significance, its decay rate does.

Given the eigenvalues (7.34), (7.36), and (7.37) one naturally defines the inverse correlation lengths:

$$\frac{1}{\xi_{\mathrm{free}}(\beta)} = -\lim_{|x|\to\infty} \frac{1}{|x|} \log G_{\mathrm{free}}(x)\,, \tag{7.38}$$

$$\frac{1}{\xi_{\mathrm{wir}}^{(1)}(\beta)} = -\lim_{|x|\to\infty} \frac{1}{|x|} \log G_{\mathrm{wir}}^{(1)}(x) \tag{7.39}$$

and

$$\frac{1}{\xi_{\mathrm{wir}}^{(2)}(\beta)} = -\lim_{|x|\to\infty} \frac{1}{|x|} \log G_{\mathrm{wir}}^{(2)}(x)\,. \tag{7.40}$$

In all cases, the limits are taken so that x lies along a coordinate axis. [BC] show that existence of the limits (7.38) and (7.40) can be established for all $q \geq 1$ by using a random cluster analogue of the subadditivity argument we discussed in Chapter 2. Unfortunately, however, the existence of the limit (7.39) does not follow from subadditivity. Instead, [BC] use a so-called reflection positivity argument in the spin representation; see the paper for more details.

All three correlation lengths coincide in the high-temperature regime, where as we mentioned earlier, their common value is often denoted by $\xi_{\mathrm{dis}}(\beta)$. In the low-temperature regime, we expect $\xi_{\mathrm{free}}(\beta) = \infty$. Also as mentioned above, the non-trivial correlation length in this regime is often denoted by $\xi_{\mathrm{ord}}(\beta)$. Here, however, we see that for $q \geq 3$, there are two *a priori* different non-trivial lengths, $\xi_{\mathrm{wir}}^{(1)}(\beta)$ and $\xi_{\mathrm{wir}}^{(2)}(\beta)$. Equations (7.36) and (7.37) immediately imply that

$$\xi_{\mathrm{wir}}^{(1)}(\beta) \,\geq\, \xi_{\mathrm{wir}}^{(2)}(\beta)\,, \tag{7.41}$$

[1] For the Ising model ($q=2$), $G_c^{mn}(x-y)$ has only the trivial eigenvalue zero and the eigenvalue $G_{\mathrm{wir}}^{(1)}(x-y)$.

so that the correlation length $\xi^{(1)}_{\text{wir}}$ of the symmetric state (i.e. symmetric with respect to S_{q-1}) is not smaller than those of the unsymmetric states. An interesting open question is whether or not the inequality is strict. It is worth noting that in percolation, analogues of $C_{\text{wir}}(x-y)$ and $\tau^{\text{fin}}_{\text{wir}}(x-y)$ in the absence of boundary conditions have equal exponential decay rates [CCGKS], which here would imply equality of $\xi^{(1)}_{\text{wir}}(\beta)$ and $\xi^{(2)}_{\text{wir}}(\beta)$. However, it is not at all clear whether the Potts models for $q \geq 3$ should have analogous behavior. In fact, motivated by the [BC], there is recent numerical evidence [JK] suggesting that the two lengths are indeed different. We consider strictness of the inequality (7.41) to be a fascinating open question.

We return finally to our original question, namely the discrepancy between the exact and numerical correlation lengths of two-dimensional Potts models with discontinuous transitions. [BC]'s resolution of the discrepancy is a relation of the conjectured form (7.30) in terms of the smaller ordered correlation length, $\xi^{(2)}_{\text{wir}}$. Their result follows from a dichotomy which they prove for all two-dimensional random cluster models with $q \geq 1$. In addition to the conjectured relation, the dichotomy implies $\nu_+ = \nu_-$ for Potts models with continuous transitions. Let $P^{\text{free}}_\infty(\beta)$ be the percolation probability in the free boundary condition random cluster measure. The [BC] dichotomy is: If $P^{\text{free}}_\infty(\beta^*) = 0$, then

$$\xi^{(2)}_{\text{wir}}(\beta) = \frac{1}{2}\xi_{\text{free}}(\beta^*)\,, \tag{7.42}$$

whereas if $P^{\text{free}}_\infty(\beta^*) > 0$, then

$$\xi_{\text{free}}(\beta) = \xi^{(1)}_{\text{wir}}(\beta) = \xi^{(2)}_{\text{wir}}(\beta)\,. \tag{7.43}$$

In order to interpret the dichotomy, we supplement it with the two-dimensional relation

$$P^{\text{wir}}_\infty(\beta)\, P^{\text{free}}_\infty(\beta^*) = 0\,, \tag{7.44}$$

where $P^{\text{wir}}_\infty(\beta)$ is the percolation probability in the wired measure, which is of course equal to the spontaneous magnetization $M(\beta)$. Note that (7.44) shows that $P^{\text{free}}_\infty(\beta^*) > 0$ implies $M(\beta) = 0$, so that (7.43) is simply the equality of the three correlation lengths in the high-temperature regime, as mentioned earlier.

The more interesting corollaries follow from the first branch of the dichotomy, i.e. the duality relation (7.42). In order to see this, we combine (7.44) with the obvious bound $P^{\text{wir}}_\infty(\beta) \geq P^{\text{free}}_\infty(\beta)$ to obtain $P^{\text{free}}_\infty(\beta)\, P^{\text{free}}_\infty(\beta^*) = 0$, so that $P^{\text{free}}_\infty(\beta_s) = 0$. Since $P^{\text{free}}_\infty(\beta^*)$ is an increasing function of β^*, this in turn implies

$$P^{\text{free}}_\infty(\beta^*) = 0 \quad \text{for all} \quad \beta \geq \beta_s\,. \tag{7.45}$$

Equation (7.45) implies in particular that $P^{\text{free}}_\infty(\beta)$ is left continuous at the self-dual point β_s. Moreover, it means that that the first branch of the dichotomy (i.e. equation (7.42)) holds throughout the low-temperature phase $\beta \geq \beta_s$. For systems with first-order transitions, this implies the conjectured relation at β_s:

$$\xi^{(2)}_{\text{wir}}(\beta_s) = \frac{1}{2}\xi_{\text{free}}(\beta_s)\,.$$

For systems with second-order transitions, (7.42) is a generalization of the aforementioned results on two-dimensional percolation [CCGKS] and the Ising magnet [MW]. In particular, it gives a strong form of Widom scaling as $\beta \to \beta_s$: If $\xi_{\text{free}}(\beta^*)$ diverges with critical exponent ν, $\xi_{\text{free}}(\beta^*) \sim |\beta^* - \beta_s|^{-\nu}$ as $\beta^* \nearrow \beta_s$, (7.42) implies that $\xi^{(2)}_{\text{wir}}(\beta)$ diverges with the same exponent: $\xi^{(2)}_{\text{wir}}(\beta) \sim |\beta - \beta_s|^{-\tilde{\nu}}$ as $\beta \searrow \beta_s$ with $\tilde{\nu} = \nu$. Moreover, the amplitudes are related by a factor of 2.

As noted above, the interpretation (and in fact, the proof) of the dichotomy (7.42) and (7.43) requires the relation (7.44), which [BC] obtain as a special case of a general two-dimensional result of Gandolfi, Keane and Russo [GKR]. However, in order to apply the [GKR] theorem, they need to know that the free random measure is ergodic, a result which [BC] establish in all dimensions. We have of course already seen most of the proof of this result in Subsection 7.6, based on the theorems of Grimmett [Grim3] and Pfister and Vande Velde [PV].

We close these notes with a heuristic, pictorial proof of the more interesting branch of the dichotomy, namely relation (7.42), under the hypothesis $P^{\text{free}}_\infty(\beta^*) = 0$. The actual proof is dozens of pages and involves technicalities which we certainly do not want to address here. Instead, we will simply indicate arguments for upper and lower bounds of the form

$$\tau^{\text{fin}}_{\text{wir}}(\beta) \lesssim (\tau_{\text{free}}(\beta^*))^2$$

and

$$\tau^{\text{fin}}_{\text{wir}}(\beta) \gtrsim (\tau_{\text{free}}(\beta^*))^2$$

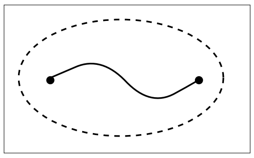
based on duality and the two decoupling inequalities. The desired relation (7.42) follows as the exponential decay rate of the resulting equality

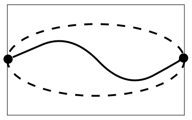
$$\tau^{\text{fin}}_{\text{wir}}(\beta)(x-y) \sim (\tau_{\text{free}}(\beta^*)(x-y))^2.$$

Here $\lesssim$, $\gtrsim$ and $\sim$ are meant in the sense of logs and limits; see [BC] for the precise statements and proofs.

The Upper Bound.

Begin by writing $\tau^{\text{fin}}_{\text{wir}}(\beta)(x-y)$ graphically and noting that the decay rate $\xi^{(2)}_{\text{wir}}(\beta)$ is the result of a double limit—first taking the volume to infinity and then taking $|x-y| \to \infty$. Here we would like to "almost interchange" these two limits, which turns out to be impossible since the desired inequality goes in the opposite direction from the *a priori* inequality. However, with a good deal of work involving the first decoupling inequality and the finite energy condition, we can essentially take these two limits simultaneously—which we will represent graphically as pulling in the boundaries:

$$\tau^{\text{fin}}_{\text{wir}}(\beta) \sim \mu^{\beta}_{\text{wir},\Lambda}\Bigg($$

$$\Bigg)$$

$$\sim \mu^{\beta}_{\text{wir},\Lambda}\Bigg($$

$$\Bigg).$$

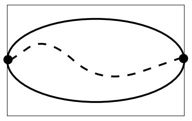
Here, of course, solid lines represent bonds and dotted lines represent dual bonds. Next we use duality, which tells us that

$$\begin{aligned} &\text{bonds} \;\to\; \text{dual bonds}, \\ &\text{dual bonds} \;\to \text{bonds}, \\ &\beta \to \beta^* \end{aligned}$$

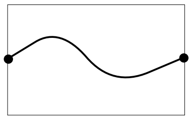
and

$$\text{wir} \;\to\; \text{free}.$$

We have

$$\tau^{\text{fin}}_{\text{wir}}(\beta) \sim \mu^{\beta^*}_{\text{free},\Lambda}\Bigg($$

$$\Bigg).$$

But this is exactly of the form needed to apply the second BC decoupling inequality. We obtain

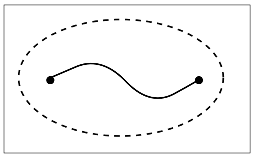
$$\tau^{\text{fin}}_{\text{wir}}(\beta) \lesssim \mu^{\beta^*}_{\text{free},\Lambda}\Bigg($$

$$\Bigg)^2$$

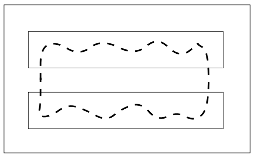
$$\sim \tau^{\text{fin}}_{\text{free}}(\beta^*)^2,$$

as desired.

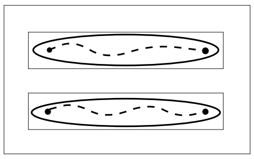
The Lower Bound.

Again we begin by writing $\tau^{\text{fin}}_{\text{wir}}(\beta)(x-y)$ graphically. For a lower bound, we note that we need only devise one mechanism to ensure that the desired event occurs. The desired events is, of course, a direct bond connection from x to y and a dual bond "bubble" around the connection. Let us require that the "bubble" be in the form of four paths: two horizontal paths, each confined to rectangles of length roughly $|x-y|$ (and width independent of $|x-y|$), and two vertical paths connecting the horizontal paths. Since the horizontal dual bond paths are confined to rectangles, it follows that each of them must be surrounded by paths of direct bonds which are the boundaries of the dual clusters in the rectangles. Thus the direct bond connection from x to y is essentially (i.e. up to factors which are uniform in $|x-y|$) ensured by the requirement that the horizontal paths stay in their rectangles. We have

$$\tau^{\mathrm{fin}}_{\mathrm{wir}}(\beta) \sim \mu^{\beta}_{\mathrm{wir},\Lambda}\Bigg($$

$$\Bigg)$$

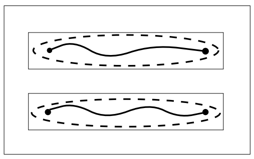
$$\gtrsim \mu^{\beta}_{\mathrm{wir},\Lambda}\Bigg($$

$$\Bigg).$$

Next, using finite energy, we remove the vertical paths (noting that the cost is uniform in $|x-y|$):

$$\tau^{\mathrm{fin}}_{\mathrm{wir}}(\beta) \gtrsim \mu^{\beta}_{\mathrm{wir},\Lambda}\Bigg($$

$$\Bigg).$$

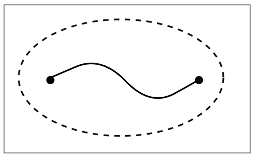
Note that we have now explicitly drawn the direct bond boundaries of the dual bond clusters.

As in the proof of the upper bound, we now use duality, which gives us

$$\tau^{\mathrm{fin}}_{\mathrm{wir}}(\beta) \gtrsim \mu^{\beta^*}_{\mathrm{free},\Lambda}\Bigg($$

$$\Bigg).$$

But this is exactly of the form needed to apply the first BC decoupling inequality. We have

$$\tau_{\mathrm{wir}}^{\mathrm{fin}}(\beta) \gtrsim \mu_{\mathrm{free},\Lambda}^{\beta^*}\Bigg($$



$$\Bigg)^2$$

$$\sim \tau_{\mathrm{free}}^{\mathrm{fin}}(\beta^*)^2.$$

This is almost what we want, except that the connections are forced to be finite, i.e. we have $\tau_{\mathrm{free}}^{\mathrm{fin}}(\beta^*)$ rather than $\tau_{\mathrm{free}}(\beta^*)$. However, we have not yet used our hypothesis $P_\infty^{\mathrm{free}}(\beta^*) = 0$. This ensures that bonds are not percolating in the free measure at temperature β^*, which means that the event we want is equal to the event we have, up to a set of measure zero. Thus $\tau_{\mathrm{free}}^{\mathrm{fin}}(\beta^*) = \tau_{\mathrm{free}}(\beta^*)$ and we have

$$\tau_{\mathrm{wir}}^{\mathrm{fin}}(\beta) \gtrsim (\tau_{\mathrm{free}}(\beta^*))^2,$$

which completes our heuristic argument.

References

[A] M. Aizenman, *Geometric analysis of ϕ^4 fields and Ising models. Parts I and II*, Commun. Math. Phys. **86** (1982), 1–48.

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

[ACC] K. Alexander, J. T. Chayes and L. Chayes, *Asymptotics of the finite cluster distribution and the Wulff construction for two-dimensional Bernoulli percolation*, Commun. Math. Phys. **131** (1990), 1–50.

[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*, Commun. Math. Phys. **92** (1983), 19–69.

[ACCN1] M. Aizenman, J. T. Chayes, L. Chayes and C. M. Newman, *The phase boundary in dilute and random Ising and Potts ferromagnets*, J. Phys. A: Math. Gen. **20** (1987), L313–L318.

[ACCN2] M. Aizenman, J. T. Chayes, L. Chayes and C. M. Newman, *Discontinuity of the magnetization in the one-dimensional $\frac{1}{|x-y|^2}$ Ising and Potts models*, J. Stat. Phys. **50** (1988), 1–40.

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

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

[AN1] M. Aizenman and C. M. Newman, *Tree graph inequalities and critical behavior in percolation models*, J. Stat. Phys. **36** (1984), 107–143.

[AN2] M. Aizenman and C. M. Newman, *Discontinuity of the percolation density in one dimensional $1/|x-y|^2$ percolation models*, Commun. Math. Phys. **107** (1986), 611–647.

[AYH] P. W. Anderson, G. Yuval and D. R. Hamann, *Exact results in the Kondo problem II. Scaling theory, qualitatively correct solution, and some new results on the one-dimensional classical statistical models*, Phys. Rev. B **1** (1970), 4464–.

[BC] C. Borgs and J. T. Chayes, *On the covariance matrix of the Potts model: A random cluster analysis*, J. Stat. Phys. **82** (1996), 1235–1297.

[BCKS] C. Borgs, J. T. Chayes, H. Kesten and J. Spencer, *The birth of the infinite cluster: finite-size scaling in percolation*, in preparation.

[Be] J. van den Berg, *private communication.*

[BF] J. van den Berg and U. Fiebig, *On a combinatorial conjecture concerning disjoint occurrences of events*, Ann. Probab. **15** (1987), 354–374.

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

[BI1] C. Borgs and J. Imbrie, *Finite-size scaling and surface tension from effective one-dimensional systems*, Commun. Math. Phys. **145** (1992), 235–280.

[BI2] C. Borgs and J. Imbrie, *Crossover finite-size scaling at first order transitions*, J. Stat. Phys. **69** (1992), 487–537.

[BJ] C. Borgs and W. Janke, *An explicit formula for the interface tension of the 2D Potts model*, J. Phys. I France **2** (1992), 2011–2018.

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

[BKM] C. Borgs, R. Kotecký and S. Miracle-Sole, *Finite-size scaling for the Potts model*, J. Stat. Phys. **62** (1991), 529–551.

[BoK] C. Borgs and R. Kotecký, *A rigorous theory of finite-size scaling at first-order phase transitions*, J. Stat. Phys. **61** (1990), 79–110.

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

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

[BW] E. Buffenoir and S. Wallon, *The correlation length of the Potts model at the first-order transition point*, Jour. Phys. A: Math. Gen. **26** (1993), 3045–3062.

[CC1] J. T. Chayes and L. Chayes, *Percolation and random media*, Random Systems and Gauge Theories, Les Houches, Session XLIII (K. Osterwalder and R. Stora, eds.), Elsevier, Amsterdam, 1984, pp. 1001-1142.

[CC2] J. T. Chayes and L. Chayes, *Critical points and intermediate phases on wedges of $\mathbb{Z}^d$*, J. Phys. A: Math. Gen. **19** (1986), 3033–3048.

[CC3] J. T. Chayes and L. Chayes, *An inequalities for the infinite cluster density in percolation*, Phys. Rev. Lett. **56** (1986), 1619–1622.

[CC4] J. T. Chayes and L. Chayes, *The mean field bound for the order parameter of Bernoulli percolation*, Percolation Theory and Ergodic Theory of Infinite Particle Systems (H. Kestin, ed.), Springer, Berlin, 1987, pp. 49–71.

[CCD] J. T. Chayes, L. Chayes and R. Durrett, *Inhomogeneous percolation problems and incipient infinite clusters*, J. Phys. A: Math. Gen. **20** (1987), 1521–1530.

[CCF] J. T. Chayes, L. Chayes and J. Fröhlich, *The low-temerature behavior of disordered magnets*, Commun. Math. Phys. **100** (1985), 399–437.

[CCFS1] J. T. Chayes, L. Chayes, D. Fisher and T. Spencer, *Finite-size scaling and correlation lengths for disordered systems*, Phys. Rev. Lett. **57** (1986), 2999-3002.

[CCFS2] J. T. Chayes, L. Chayes, D. S. Fisher and T. Spencer,, *Correlation length bounds for disordered Ising ferromagnets*, Commun. Math. Phys. **120** (1989), 501–523.

[CCGKS] J. T. Chayes, L. Chayes, G. R. Grimmett, H. Kesten and R. H. Schonmann, *The correlation length for the high-density phase of Bernoulli percolation*, Ann. Probab. **17** (1989), 1277–1302.

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

[CCN2] J. T. Chayes, L. Chayes and C. M. Newman, *Bernoulli percolation above threshold: an invasion percolation analysis*, Ann. Probab. **15** (1987), 1272–1287.

[Do1] R. L. Dobrushin, *Gibbsian random fields for lattice systems with pairwise interactions*, Funct. Anal. Appl. **2** (1968), 292–301.

[Do2] R. L. Dobrushin, *The description of a random field by means of conditional probabilities and conditions of its regularity*, Theory Probab. Appl. **13** (1968), 197–224.

[Dy1] F. J. Dyson, *Existence of a phase transition in a one-dimensional Ising ferromagnet*, Commun. Math. Phys. **12** (1969), 91–107.

[Dy2] F. J. Dyson, *An Ising ferromanget with discontinuous long-range order*, Commun. Math. Phys. **21** (1971), 269–283.

[ER1] P. Erdös and A. Rényi, *On random graphs I*, Publ. Math. (Debrecen) **6** (1959), 290-297.

[ER2] P. Erdös and A. Rényi, *On the evolution of random graphs*, Magy. Tud. Akad. Mat. Kut. Intéz. Közl. **5** (1960), 17–61.

[ES] R. G. Edwards and A. D. Sokal, *Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm*, Phys. Rev. D **38** (1988), 2009–2012.

[FiS] P.C. Fishburn and L.A. Shepp, *On the FKB conjecture for disjoint intersections*, Discrete Math. **98** (1991), 105–122.

[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.

[FKG] C. M. Fortuin, P. W. Kasteleyn and J. Ginibre, *Correlation inequalities on some partially ordered sets*, Commun. Math. Phys. **22** (1971), 89–103.

[Fo] C. M. Fortuin, *On the random cluster model. III. The simple random-cluster process*, Physica **59** (1972), 545–570.

[Fr] J. Fröhlich, *On the triviality of $\lambda\varphi_d^4$ theories and the approach to the critical point in $d \geq 4$ dimensions*, Nucl. Phys. B **200** (1982), 281–296.

[FrK] E. Friedgut and G. Kalai, *Every monotone graph property has a sharp threshold*, preprint (1996).

[FrS] J. Fröhlich and T. Spencer, *The phase transition in the one-dimensional Ising model with $1/r^2$ interaction energy*, Commun. Math. Phys. **84** (1982), 87–101.

[Ga] A. Gandolfi, *Uniqueness of the infinite cluster for stationary Gibbs states*, Ann. Probab. **17** (1989), 1403–1415.

[Ge] H-O. Georgii, *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin, New York, 1988.

[GGR] A. Gandolfi, G. Grimmett and L. Russo, *On the uniqueness of the infinite cluster in the percolation model*, Commun. Math. Phys. **114** (1988), 549–552.

[GI] S. Gupta and A. Irbäck, *Physics beyond instantons: Measuring the physical correlation length*, Phys. Lett. B **286** (1992), 112–117.

[Gin] J. Ginibre, *General formulation of Griffiths' inequalities*, Commun. Math. Phys. **16** (1970), 310–328.

[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*, Probab. Theory Relat. Fields **92**, 511-527.

[GKR] A. Gandolfi, M. Keane and L. Russo, *On the uniqueness of the infinite occupied cluster in dependent two-dimensional site percolation*, Ann. Probab. **16** (1988), 1147-1157.

[Grif] R. B. Griffiths, *Correlations in Ising ferromagnets. II. External magnetic fields*, Jour. Math. Phys. **8** (1967), 484–489.

[Grim1] G. R. Grimmett, *Bond percolation on subsets of the square lattice, and the transition between one-dimensional and two-dimensional behavior*, J. Phys. A: Math. Gen. **16** (1983), 599–604.

[Grim2] G. Grimmett, *Percolation*, Springer-Verlag, New York, 1989.

[Grim3] G. Grimmett, *The stochastic random-cluster process and the uniqueness of random-cluster measures*, preprint (1994).

[Ham1] J. M. Hammersley, *Percolation processes. Lower bounds for the critical probability*, Ann. Math. Stat. **28** (1957), 790–795.

[Ham2] J. M. Hammersley, *in Methods in Computational Physics, Volume I* (B. Adler, S. Fernbach and M. Rotenberg, eds.), Academic Press, London, 1963, pp. 281–298.

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

[HS] T. Hara and G. Slade, *Mean-field critical behavior for percolation in high dimensions*, Commun. Math. Phys. **128** (1990), 333-391.

[HW] J. M. Hammersley and S. G. Whittington, *Self-avoiding walks in wedges*, J. Phys. A: Math. Gen. **18** (1985), 101–111.

[IN] J. Imbrie and C. M. Newman, *An intermediate phase with slow decay of correlations in one dimensional* $1/|x-y|^2$ *percolation, Ising and Potts models*, Commun. Math. Phys. **118** (1988), 303–336.

[JK] W. Janke and S. Kappler, *2D Potts model correlation lengths: Numerical evidence for* $\xi_o = \xi_d$ *at* β_t, Europhys. Lett. **31** (1995), 345–.

[JKLP] S. Janson, D. E. Knuth, T. Luczak and B. Pittel, *The birth of the giant component*, Random Struc. Alg. **4** (1993), 233–358.

[K1] H. Kesten, *The critical probability of bond percolation on the square lattice equals* 1/2, Commun. Math. Phys. **74** (1980), 41–59.

[K2] H. Kesten, *Percolation Theory for Mathematicians*, Birkhäuser, Boston, 1982.

[K3] H. Kesten, *The incipient infinite cluster in two-dimensional percolation*, Probab. Th. Rel. Fields **73** (1986), 369–394.

[K4] H. Kesten, *Scaling relations for 2D-percolation*, Commun. Math. Phys. **109** (1987), 109–156.

[Kl] A. Klümper, *Investigation of excitation spectra of exactly solvable models using inversion relations*, Int. J. Mod. Phys. B **4** (1990), 871–893.

[KoS] R. Kotecký and S. Shlosman, *First-order phase transitions in large entropy lattice systems*, Commun. Math. Phys. **83** (1982), 493-550.

[KSZ] A. Klümper, A. Schadschneider and J. Zittartz, *Inversion relations, phase transitions and transfer matrix excitations for special spin models in two dimensions*, Z. Phys. B **76** (1989), 247–258.

[KZ] H. Kesten and Y. Zhang, *The probability of a large finite cluster in supercritical Bernoulli percolation*, Ann. Probab. **18** (1990), 537–555.

[L] E. H. Lieb, *A refinement of Simon's correlation inequality*, Commun. Math. Phys. **77** (1980), 127–135.

[LMR] L. Laanait, A. Messager and J. Ruiz, *Phase coexistence and surface tensions for the Potts model*, Commun. Math. Phys. **105** (1986), 527–545.

[LMMRS] L. Laanait, A. Messager, S. Miracle-Sole, J. Ruiz and S. Shlosman, *Interfaces in the Potts model I: Pirogov-Sinai theory of the Fortuin-Kasteleyn representation*, Commun. Math. Phys. **140** (1991), 81–91.

[LR] O. E. Lanford and D. Ruelle, *Observables at infinity and states with short range correlations in statistical mechanics*, Commun. Math. Phys. **13**, 194-215.

[Ma] G. A. Margulis, *Probabilistic characteristics of graphs with large connectivity*, Prob. Peredachi Inform. **10** (1977), 101–108.

[MMRS] A. Messager, S. Miracle-Sole, J. Ruiz and S. Shlosman, *Interfaces in the Potts model II: Antonov's rule and rigidity of the order disorder interface*, Commun. Math. Phys. **140** (1991), 275–290.

[MMS] M. V. Menshikov, S. A. Molchanov and A. F. Sidorenko, *Percolation theory and some applications*, Itogi Nauki i Techniki (Series of Probability Theory, Mathematical Statistics, Theoretical Cybernetics) **24** (1986), 53–110.

[MS] N. Madras and G. Slade, *The Self-Avoiding Walk*, Birkhäuser, Boston, 1993.

[MW] B. McCoy and T. T. Wu, *The Two-Dimensional Ising Model*, Harvard Univ. Press, Cambridge, Mass, 1973.

[NS1] C. M. Newman and L. S. Schulman, *Infinite clusters in percolation models*, Jour. Stat. Phys. **26** (1981), 613–628.

[NS2] C. M. Newman and L. S. Schulman, *One-dimensional $1/|j-i|^s$ percolation models: The existence of a transition for $s \leq 2$*, Commun. Math. Phys. **104** (1986), 547–571.

[PL] P. Peczak and D. P. Landau, *Monte Carlo study of finite-size effects at a weakly first-order phase transition*, Phys. Rev. B **39** (1989), 11932–11942.

[PV] C. E. Pfister and K. Vande Velde, *Almost sure quasilocality in the random cluster model*, J. Stat. Phys. **79** (1995), 765–774.

[Pr] C. Preston, *Random Fields*, Springer-Verlag, Berlin, Heidelberg, New York, 1976.

[R1] L. Russo, *A note on percolation*, Z. Wahrsch. Geb. **43** (1978), 39–48.

[R2] L. Russo, *On the critical percolation probabilities*, Z. Wahrsch. Geb. **56** (1981), 229–237.

[Re] D. Reimer, *preliminary manuscript.*

[Rue] D. Ruelle, *Statistical mechanics of a one-dimensional lattice gas*, Commun. Math. Phys. **9** (1968), 267–278.

[S] B. Simon, *Correlation inequalities and the decay of correlations in ferromagnets*, Commun. Math. Phys. **77** (1980), 111–126.

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

[SW] R. H. Swendsen and J. S. Wang, *Nonuniversal critical dynamics in Monte Carlo simulations*, Phys. Rev. Lett. **58** (1987), 86–88.

[Ta] M. Talagrand, *Some remarks on the Berg-Kesten inequality, Probab. in Banach Spaces, 9 (Sandjberg, 1993)*, Prog. Probab. **35** (1994), Birkhäuser, Boston, 293–297.

[Th] D. J. Thouless, *Long-range order in one-dimensional Ising systems*, Phys. Rev. **187** (1969), 732–733.

[W] F. Y. Wu, *Knot theory and statistical mechanics*, Rev. Mod. Phys. **64** (1992), 1099–1131.