

collaborating on various enjoyable and fascinating projects. GS expresses special thanks to David Brydges for inspiring his initial interest in the lace expansion and suggesting that it might converge in five dimensions, and to Takashi Hara for the pleasure of four years of collaboration and for permission to include unpublished joint work. We gratefully acknowledge financial support from the Natural Sciences and Engineering Research Council of Canada. Finally we offer our thanks and deep appreciation to Joyce Kruskal and Joanne Nakonechny for their encouragement, support, and tolerance.

Toronto and Hamilton
July 7, 1992

Chapter 1

Introduction

1.1 The basic questions

Imagine that you are standing at an intersection in the centre of a large city whose streets are laid out in a square grid. You choose a street at random and begin walking away from your starting point, and at each intersection you reach you choose to continue straight ahead or to turn left or right. There is only one rule: you must not return to any intersection already visited in your journey. In other words, your path should be self-avoiding. It is possible that you will lead yourself into a trap, reaching an intersection whose neighbours have all been visited already, but barring this disaster you continue walking until you have walked some large number N of blocks. There are two basic questions:

- How many possible paths could you have followed?
- Assuming that any one path is just as likely as any other, how far will you be on the average from your starting point?

These questions are straightforward enough, but the answers are only known for small values of N . It is widely accepted that a search for general exact formulas is an enormously difficult problem which lies beyond the reach of current methods. A less difficult question would be to ask for the asymptotic behaviour of the answers as N becomes very large, but this too is very hard. Physicists and chemists who are interested in this and related problems have applied a variety of methods and have produced many intriguing results, but a great deal of work is still needed to settle these issues in a mathematically rigorous way. In this book we will state some of the

results of nonrigorous work in the field, and describe the rigorous work in some detail.

At first glance one might expect that the easiest way to answer the above questions, at least approximately, would be to use a computer. Much numerical work has been done in this direction, and in Chapter 9 some of it will be discussed. Here too, however, the situation is not so easy: exact enumeration of all possible routes has been done to date only for $N \leq 34$, with further enumerations made difficult because of the exponential growth in the number of paths as N increases. Larger values of N can be studied by extrapolation of the exact enumeration data, or by Monte Carlo simulations.

There is no need to restrict the walk to a two-dimensional grid, and it is easy to generalize the above questions to general dimension d . It is also possible to generalize the problem by changing from a rectangular to a triangular or other type of grid. There is at least one case where the above questions can be easily answered, and this is the case of a one-dimensional walk. A self-avoiding walker in one dimension has no alternative but to continue travelling in the direction initially chosen, so there are exactly two paths for every value of N and the distance travelled is exactly N blocks. That was easy, but not very interesting. Higher dimensions provide a vastly richer structure.

In general, a self-avoiding walk takes place on a graph. A graph (more precisely, an undirected graph) is a collection of points, together with a collection of pairs of points known as *edges*. The basic example that will concern us most is the d -dimensional hypercubic lattice $\mathbb{Z}^d$. The points of this graph are the points of the d -dimensional Euclidean space $\mathbb{R}^d$ whose components are all integers, and the edges are given by the set of all unit line segments joining neighbouring points. The points will be referred to as *sites*, and the unit line segments as *nearest-neighbour bonds*. Sites will typically be denoted by letters such as u, v, x, y , and their components by subscripts: $x = (x_1, x_2, \dots, x_d)$. The usual Euclidean dot product on $\mathbb{Z}^d$ will be written $x \cdot y = \sum_{i=1}^d x_i y_i$, and the Euclidean norm will be written $|x| = \sqrt{x \cdot x}$. We will also use the notation $\|x\|_p = (\sum_{i=1}^d x_i^p)^{1/p}$, and $\|x\|_\infty = \max\{|x_i|: i = 1, \dots, d\}$.

An N -step self-avoiding walk ω on $\mathbb{Z}^d$, beginning at the site x , is defined as a sequence of sites $(\omega(0), \omega(1), \dots, \omega(N))$ with $\omega(0) = x$, satisfying $|\omega(j+1) - \omega(j)| = 1$, and $\omega(i) \neq \omega(j)$ for all $i \neq j$. We write $|\omega| = N$ to denote the length of ω , and we denote the components of $\omega(j)$ by $\omega_i(j)$ ($i = 1, \dots, d$). Let c_N denote the number of N -step self-avoiding walks beginning at the origin. By convention, $c_0 = 1$. Then the first of our basic questions above is asking for the value of c_N . More modestly, we could ask

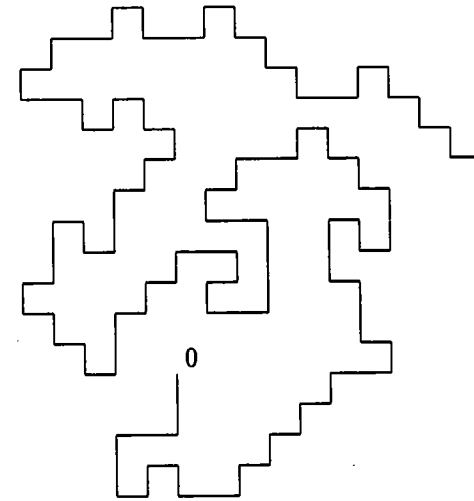


Figure 1.1: A two-dimensional self-avoiding walk with 115 steps.

for the asymptotic form of c_N as $N \rightarrow \infty$. It is easy to find the exact values of c_N (as a function of d) for very small values of N , for example $c_1 = 2d$, $c_2 = 2d(2d-1)$, $c_3 = 2d(2d-1)^2$, and $c_4 = 2d(2d-1)^3 - 2d(2d-2)$ (for c_4 the second term subtracts the contribution of squares to the first term). However, the combinatorics quickly become difficult as N increases and then soon become intractable. Tables in Appendix C give enumerations of c_N for dimensions two through six.

The simplest bounds on the behaviour of c_N are obtained as follows. An upper bound on c_N is given by the number of walks which have no immediate reversals, or in other words which never visit the same site at times i and $i+2$. Avoiding immediate reversals allows $2d$ choices for the initial step, and $2d-1$ choices for the $N-1$ remaining steps, for a total of $2d(2d-1)^{N-1}$. For a lower bound we simply count the number of walks in which each step is in one of the d positive coordinate directions. Such walks are necessarily self-avoiding. Thus we have

$$d^N \leq c_N \leq 2d(2d-1)^{N-1}. \quad (1.1.1)$$

To discuss the average distance from the origin after N steps, we need to introduce a probability measure on N -step self-avoiding walks. The measure that we shall use throughout this book is the uniform measure, which assigns equal weight c_N^{-1} to each N -step self-avoiding walk. It is worth noting that although we originally introduced the self-avoiding walk

in terms of a walker moving in time, the uniform measure is a measure on paths of length N and does not define a stochastic process evolving in time (for example, a walk may be trapped and impossible to extend without introducing a self-intersection).

Denoting expectation with respect to the uniform measure by angular brackets, the average distance (squared) from the origin after N steps is then given by the *mean-square displacement*

$$\langle |\omega(N)|^2 \rangle = \frac{1}{c_N} \sum_{\omega: |\omega|=N} |\omega(N)|^2. \quad (1.1.2)$$

The sum over ω is the sum over all N -step self-avoiding walks beginning at the origin. Like c_N , the mean-square displacement can also be calculated by hand for very small values of N , but the combinatorics quickly become intractable as N increases. Enumerations are tabulated in Appendix C.

It is instructive to compare the behaviour of the self-avoiding walk with that of the simple random walk. An N -step simple random walk on $\mathbb{Z}^d$, starting at the origin, is a sequence $\omega = (\omega(0), \omega(1), \dots, \omega(N))$ of sites with $\omega(0) = 0$ and $|\omega(j+1) - \omega(j)| = 1$, with the uniform measure on the set of all such walks. Without the self-avoidance constraint the situation is rather easy. Indeed, since each site has $2d$ nearest neighbours, the number of N -step simple random walks is exactly $(2d)^N$. To analyse the mean-square displacement, we represent the simple random walk in the following way. Let $\{X^{(i)}\}$ be independent and identically distributed random variables with $X^{(i)}$ uniformly distributed over the $2d$ (positive and negative) unit vectors. Then the position after N steps can be represented as the sum $S_N = X^{(1)} + X^{(2)} + \dots + X^{(N)}$. Expanding $|S_N|^2$, the mean-square displacement is given by

$$\langle |S_N|^2 \rangle = \sum_{i,j=1}^N \langle X^{(i)} \cdot X^{(j)} \rangle. \quad (1.1.3)$$

For $i \neq j$, $\langle X^{(i)} \cdot X^{(j)} \rangle = 0$, using independence and the fact that $\langle X^{(i)} \rangle = 0$. Since $\langle X^{(i)} \cdot X^{(i)} \rangle = 1$, it follows that the mean-square displacement is equal to N . Similarly, if we consider a random walk in $\mathbb{Z}^d$ in which steps lie in a symmetric finite set $\Omega \subset \mathbb{Z}^d$ of cardinality $|\Omega|$, with each possible step equally likely, then the number of N -step walks is $|\Omega|^N$ and the mean-square displacement is $N\sigma^2$, where σ^2 is the mean-square displacement of a single step.

For the self-avoiding walk it is believed that there is exponential growth of c_N with power law corrections, unlike the pure exponential growth of

the simple random walk. It is also believed that the mean-square displacement will not always be linear in the number of steps, in contrast to the diffusive behaviour of the simple random walk. These beliefs are in harmony with known properties of other models of statistical mechanics, and are supported by numerical and nonrigorous calculations. The conjectured behaviour of c_N and $\langle |\omega(N)|^2 \rangle$ is thus

$$c_N \sim A\mu^N N^{\gamma-1} \quad (1.1.4)$$

and

$$\langle |\omega(N)|^2 \rangle \sim DN^{2\nu}, \quad (1.1.5)$$

where A , D , μ , γ and ν are dimension-dependent positive constants. We shall refer to μ as the *connective constant*, and γ and ν are examples of *critical exponents*. In four dimensions the above two relations should be modified by logarithmic factors; see (1.1.13) and (1.1.14) below. Here $f(N) \sim g(N)$ means that f is asymptotic to g as $N \rightarrow \infty$:

$$\lim_{N \rightarrow \infty} \frac{f(N)}{g(N)} = 1.$$

For ordinary random walk (1.1.4) and (1.1.5) hold with $\gamma = 1$ and $\nu = 1/2$, both for the nearest-neighbour and more general walks.

In the next section the existence of the limit

$$\mu = \lim_{N \rightarrow \infty} c_N^{1/N} \quad (1.1.6)$$

will be proven, which is the first step in justifying (1.1.4). The simple bounds of (1.1.1) then immediately imply that

$$d \leq \mu \leq 2d - 1. \quad (1.1.7)$$

The exact value of μ is not known for the hypercubic lattice in any dimension $d \geq 2$, although for the honeycomb lattice in two dimensions there is nonrigorous evidence that $\mu = \sqrt{2} + \sqrt{2}$. Improvements to (1.1.7) will be discussed in the next section. For high dimensions it is known that as $d \rightarrow \infty$

$$\mu = 2d - 1 - \frac{1}{2d} - \frac{3}{(2d)^2} + O\left(\frac{1}{(2d)^3}\right); \quad (1.1.8)$$

references are given in the Notes. In fact Fisher and Sykes (1959) established the coefficients in the $1/d$ expansion up to and including order d^{-4} , although there is no rigorous control of their error term. Intuitively (1.1.8) says that in high dimensions the principal effect of the self-avoidance constraint is to rule out immediate reversals.

Concerning γ , we will show in Section 1.2 that $c_N \geq \mu^N$ and hence $\gamma \geq 1$ in all dimensions. There is still no proof, however, that γ is finite in two, three or four dimensions, where the best bounds are

$$c_N \leq \begin{cases} \mu^N \exp[KN^{1/2}] & d = 2 \\ \mu^N \exp[KN^{2/(2+d)} \log N] & d = 3, 4 \end{cases} \quad (1.1.9)$$

for a positive constant K ; these bounds will be discussed in Sections 3.1 and 3.3. In Chapter 6 we will describe a proof that (1.1.4) holds with $\gamma = 1$ for $d \geq 5$. In addition to characterizing the asymptotic behaviour of c_N , the exponent γ provides a measure of the probability that two N -step self-avoiding walks starting at the same point do not intersect. In fact, this probability is equal to c_{2N}/c_N^2 , and assuming (1.1.4) we have

$$\frac{c_{2N}}{c_N^2} \sim \frac{2^{\gamma-1}}{A} N^{1-\gamma}. \quad (1.1.10)$$

If $\gamma > 1$ then this probability goes to zero as $N \rightarrow \infty$, while if $\gamma = 1$ it remains positive. For the simple random walk the analogous probability is known to remain positive as $N \rightarrow \infty$ for $d > 4$, and roughly speaking to go to zero like $(\log N)^{-1/2}$ for $d = 4$ and as an inverse power of N for $d = 2, 3$. A survey of the simple random walk results is given in Section 10.3.

Intuitively it is to be expected that the repulsive interaction of the self-avoiding walk will tend to drive the endpoint of the walk away from the origin faster than for simple random walk, or in other words that $\nu \geq 1/2$. However it is still an open question to prove that this “obvious” inequality $\langle |\omega(N)|^2 \rangle \geq CN$ holds in all dimensions. On the other hand, bounding $\langle |\omega(N)|^2 \rangle$ above by N^2 in (1.1.2) gives the upper bound $\langle |\omega(N)|^2 \rangle \leq N^2$, or $\nu \leq 1$. This bound is optimal in one dimension, but seems far from optimal in two or more dimensions. No upper bound of the form $CN^{2-\epsilon}$ ($C, \epsilon > 0$), or in other words $\nu < 1$, has been proven for dimensions two, three or four, however. For $d \geq 5$ it has been proved that $\nu = 1/2$; this proof will be described in Chapter 6. It will also be shown that for high dimensions the diffusion constant D is strictly greater than the simple random walk value of 1. Thus in high dimensions the self-avoiding walk does move away from the origin more quickly than the simple random walk, but only at the level of the diffusion constant and not at the level of the exponent ν . The tendency of the self-avoiding walk to move away from the origin more quickly than the simple random walk should become less pronounced as the dimension increases, and hence it is to be expected that ν is a nonincreasing function of the dimension.

The critical exponents γ and ν are believed to be dimension dependent, but independent of the type of allowed steps (as long as there are only

finitely many possible steps and the allowed steps are symmetric) or even of the type of lattice—the exponents are believed, for example, to be the same for the square and triangular lattices. This lack of dependence on the detailed definition of the model is known as *universality*, and models with the same exponents are said to be in the same *universality class*. The *connective constant* μ appearing in (1.1.4) represents the effective coordination number of the lattice and is not universal—it depends on the details of the allowed steps and the underlying lattice, as well as the dimension d .

It seems clear that in high dimensions the self-avoiding walk should be closer to the simple random walk than in low dimensions, since a simple random walk is less likely to intersect itself in high dimensions. Four dimensions plays a special role: for simple random walk the expected time of the first return to the origin, conditioned on the event that this return occurs, is finite for $d > 4$; this suggests that above four dimensions self-avoidance is a short-range effect rather than a long-range one, and hence that it will not affect the critical exponents. In addition, as mentioned above, the probability that two independent simple random walks of length N do not intersect remains bounded away from zero as $N \rightarrow \infty$ for $d > 4$, but not for $d \leq 4$.

The conjectured values of γ and ν are as follows:

$$\gamma = \begin{cases} \frac{43}{32} & d = 2 \\ 1.162... & d = 3 \\ 1 \text{ with logarithmic corrections} & d = 4 \\ 1 & d \geq 5 \end{cases} \quad (1.1.11)$$

$$\nu = \begin{cases} \frac{3}{4} & d = 2 \\ 0.59... & d = 3 \\ \frac{1}{2} \text{ with logarithmic corrections} & d = 4 \\ \frac{1}{2} & d \geq 5 \end{cases} \quad (1.1.12)$$

Currently the only rigorous results which prove power law behaviour and confirm the conjectured values of γ and ν are for $d \geq 5$. These are discussed in detail in Chapter 6. The conjectured logarithmic corrections to γ and ν in four dimensions, predicted by the renormalization group, are given by:

$$c_N \sim A\mu^N [\log N]^{1/4}, \quad d = 4 \quad (1.1.13)$$

$$\langle |\omega(N)|^2 \rangle \sim DN [\log N]^{1/4}, \quad d = 4. \quad (1.1.14)$$

Equations (1.1.11) to (1.1.14) are typical of what is found for other statistical mechanical models, such as the Ising model or percolation. A common feature is the existence of a certain dimension, the so-called *upper critical*

dimension, at which there are logarithmic corrections to critical exponents and above which all critical exponents are dimension independent and are given by the corresponding critical exponents for a simpler model, known as the *mean-field*¹ model. For the self-avoiding walk the mean-field model is the simple random walk and the simple random walk critical exponents are sometimes referred to as the mean-field exponents.

The rational values for two dimensions given in (1.1.11) and (1.1.12) come from a nonrigorous exact solution of the $O(N)$ spin model which includes the self-avoiding walk as the special case $N = 0$ (see Section 2.3). This remarkable work exploits a connection between the $O(N)$ model and the Coulomb gas and uses the renormalization group. From a different approach, nonrigorous conformal invariance arguments reproduce the same rational values. There is no analogous exact solution in three dimensions, and the $d = 3$ values given in (1.1.11) and (1.1.12) are from numerical results and field-theoretic calculations using the ϵ -expansion. References for these topics are given in the Notes.

An early conjecture for the values of ν was made by Flory, and will be discussed in Section 2.2. The Flory exponents are given by $\nu_{\text{Flory}} = 3/(2+d)$ for $d \leq 4$ and $\nu_{\text{Flory}} = 1/2$ for $d > 4$. This agrees with Equation (1.1.12) for $d = 2$ and $d \geq 4$ (apart from the logarithmic correction when $d = 4$), and comes very close for $d = 3$. The exact Flory value $\nu_{\text{Flory}} = 3/5$ in three dimensions has been ruled out by numerical work, however.

1.2 The connective constant

If (1.1.4) correctly represents the behaviour of c_N for large N , then the limit

$$\mu = \lim_{N \rightarrow \infty} c_N^{1/N} \quad (1.2.1)$$

must exist. One purpose of this section is to prove the existence of this limit as a simple consequence of a subadditive property of $\log c_N$. It then follows immediately from (1.1.1) that

$$d \leq \mu \leq 2d - 1. \quad (1.2.2)$$

The proof involves the notion of concatenation of two self-avoiding walks.

¹This terminology has its origin in the Ising model. For the Ising model the upper critical dimension is also four, and above four dimensions critical exponents are given by the exactly solvable model in which a spin interacts with the *average* of all the other spins. References are given in the Notes.

1.2. THE CONNECTIVE CONSTANT

Definition 1.2.1 The concatenation $\omega^{(1)} \circ \omega^{(2)}$ of an M -step self-avoiding walk $\omega^{(2)}$ to an N -step self-avoiding walk $\omega^{(1)}$ is the $(N+M)$ -step walk ω , which in general need not be self-avoiding, given by

$$\begin{aligned} \omega(k) &= \omega^{(1)}(k), & k &= 0, \dots, N \\ \omega(k) &= \omega^{(1)}(N) + \omega^{(2)}(k - N) - \omega^{(2)}(0), & k &= N+1, \dots, N+M. \end{aligned}$$

The product $c_N c_M$ is equal to the cardinality of the set of $(N+M)$ -step simple random walks which are self-avoiding for the initial N steps and the final M steps, but which may not be completely self-avoiding. This can be seen by concatenations of M -step walks to N -step walks, and implies that

$$c_{N+M} \leq c_N c_M. \quad (1.2.3)$$

In fact equality holds in (1.2.3) only if N or M is zero, since otherwise there will be at least one M -step walk whose concatenation with a given N -step walk fails to be self-avoiding. Taking logarithms in (1.2.3) shows that the sequence $\{\log c_n\}$ is *subadditive*:

$$\log c_{N+M} \leq \log c_N + \log c_M. \quad (1.2.4)$$

The existence of the limit (1.2.1) is a consequence of (1.2.4) and the following standard result; this was first observed by Hammersley and Morton (1954).

Lemma 1.2.2 Let $\{a_n\}_{n \geq 1}$ be a sequence of real numbers which is subadditive, i.e., $a_{n+m} \leq a_n + a_m$. Then the limit $\lim_{n \rightarrow \infty} n^{-1} a_n$ exists in $[-\infty, \infty)$ and is equal to

$$\lim_{n \rightarrow \infty} \frac{a_n}{n} = \inf_{n \geq 1} \frac{a_n}{n}. \quad (1.2.5)$$

Proof. It suffices to show that

$$\limsup_{n \rightarrow \infty} \frac{a_n}{n} \leq \frac{a_k}{k} \quad (1.2.6)$$

for every k , since taking the $\liminf_{k \rightarrow \infty}$ in (1.2.6) gives existence of the limit, and then (1.2.5) can be seen by taking the $\inf_{k \geq 1}$ in (1.2.6).

To prove (1.2.6), we fix k and let

$$A_k = \max_{1 \leq r \leq k} a_r. \quad (1.2.7)$$

Given a positive integer n we let j denote the largest integer which is strictly less than n/k . Then $n = jk + r$ for some integer r with $1 \leq r \leq k$. Using subadditivity, we have

$$a_n \leq ja_k + a_r \leq \frac{n}{k} a_k + A_k. \quad (1.2.8)$$

Dividing by n and taking the $\limsup_{n \rightarrow \infty}$ then gives (1.2.6).

Equation (1.2.5) shows that $\lim_{n \rightarrow \infty} n^{-1}a_n < \infty$. In general, the possibility that the limit equals $-\infty$ cannot be excluded, as is illustrated by the example of $a_n = -n^2$. For many applications, however, this is ruled out by an *a priori* bound such as $a_n \geq 0$. $\square$

Together with (1.2.4), Lemma 1.2.2 implies the existence of the limit $\log \mu \equiv \lim_{N \rightarrow \infty} N^{-1} \log c_N$, and hence gives (1.2.1). In fact (1.2.5) shows more:

$$\log \mu = \inf_{N \geq 1} N^{-1} \log c_N, \quad (1.2.9)$$

and hence

$$\mu^N \leq c_N, \quad N \geq 1. \quad (1.2.10)$$

This inequality can be summarized by the statement $\gamma \geq 1$, where γ is as introduced in (1.1.4), although strictly speaking we do not know that γ exists. Equation (1.2.10) also yields $\mu \leq c_N^{1/N}$. This gives a sequence of upper bounds for μ , but they converge to μ very slowly. A better bound is

$$\mu \leq \left(\frac{c_N}{c_1} \right)^{1/(N-1)}, \quad N \geq 2. \quad (1.2.11)$$

References for this and other improvements are given in the Notes.

Another sequence of upper bounds for μ can be obtained by considering walks which are self-avoiding only over a finite time scale or *memory* τ . We define $c_{N,\tau}$ to be the number of N -step walks ω beginning at the origin, for which $\omega(i) \neq \omega(j)$ whenever $0 < |i - j| \leq \tau$. Self-intersections occurring after an interval of more than τ steps are permitted. For example, $c_{N,2} = 2d(2d-1)^{N-1}$ for $N \geq 1$, since memory $\tau = 2$ simply rules out immediate reversals. For $\tau \geq N$, $c_{N,\tau} = c_N$. Memory $\tau = 0$ corresponds to the simple random walk.

The sequence $\{\log c_{N,\tau}\}_{N=1}^{\infty}$ is subadditive for every τ (for the same reason that $\{\log c_N\}_{N=1}^{\infty}$ is), and hence by Lemma 1.2.2 there is a μ_τ such that

$$\mu_\tau = \lim_{N \rightarrow \infty} c_{N,\tau}^{1/N} = \inf_{N \geq 1} c_{N,\tau}^{1/N}. \quad (1.2.12)$$

Since $c_{N,\tau} \geq c_N$, μ_τ provides an upper bound for μ . The next lemma shows that this sequence of upper bounds converges monotonically to μ .

Lemma 1.2.3 $\mu_\tau \searrow \mu$ as $\tau \rightarrow \infty$.

Proof. For $\sigma \leq \tau$, $c_{N,\sigma} \geq c_{N,\tau}$ and hence $\mu_\sigma \geq \mu_\tau$. By (1.2.12), $\mu_\tau \leq c_{N,\tau}^{1/N}$ for all N, τ . Taking $N = \tau$ gives

$$\mu \leq \mu_\tau \leq c_{\tau,\tau}^{1/\tau} = c_\tau^{1/\tau}. \quad (1.2.13)$$

Taking the limit $\tau \rightarrow \infty$ and using (1.2.1) gives the desired result. $\square$

The connective constant for the walk with memory $\tau = 4$ was shown in Fisher and Sykes (1959) to be given by the largest root of the cubic equation

$$\theta^3 - 2(d-1)\theta^2 - 2(d-1)\theta - 1 = 0. \quad (1.2.14)$$

For $d = 2$ this gives $\mu_4(2) = 2.8312$, where we have made the dimension dependence explicit by writing $\mu_\tau(d)$.

A number of investigations into the self-avoiding walk have approached the problem via the limit of finite memory walks as the memory goes to infinity. This approach was used in particular by Brydges and Spencer (1985) in applying their lace expansion to study weakly self-avoiding walk for $d > 4$, and will be adopted in Section 6.8 to obtain an upper bound in high dimensions on $c_N(0, x)$, the number of N -step self-avoiding walks which begin at the origin and end at x .

A lower bound on μ can be obtained in terms of *bridges*.

Definition 1.2.4 An N -step bridge is defined to be an N -step self-avoiding walk ω whose first components satisfy the inequality

$$\omega_1(0) < \omega_1(i) \leq \omega_1(N)$$

for $1 \leq i \leq N$. The number of N -step bridges starting at the origin is denoted b_N . By convention, $b_0 = 1$.

The concatenation of two bridges will always yield another bridge, so

$$b_M b_N \leq b_{M+N}. \quad (1.2.15)$$

Hence $\{-\log b_n\}$ is subadditive and so by Lemma 1.2.2 the limit

$$\mu_{\text{Bridge}} \equiv \lim_{n \rightarrow \infty} b_n^{1/n} = \sup_{n \geq 1} b_n^{1/n} \quad (1.2.16)$$

exists. Clearly $b_n \leq c_n$. Therefore $\mu_{\text{Bridge}} \leq \mu$, and so by (1.2.16)

$$b_N^{1/N} \leq \mu_{\text{Bridge}} \leq \mu. \quad (1.2.17)$$

In Section 3.1 it will be shown that in fact $\mu_{\text{Bridge}} = \mu$. Although the lower bound (1.2.17) is very slowly convergent, a more sophisticated use of bridges leads to better lower bounds. References can be found in the Notes at the end of this chapter.

We conclude this section with a table showing the current best rigorous upper and lower bounds on μ , together with estimates of the precise value, for the hypercubic lattice in dimensions $d = 2, 3, 4, 5, 6$.

d	lower bound	estimate	upper bound
2	2.61987 ^a	$2.6381585 \pm 0.0000010^d$	2.69576 ^b
3	4.43733 ^c	4.6839066 ± 0.0002^e	4.756 ^b
4	6.71800 ^c	6.7720 ± 0.0005^f	6.832 ^b
5	8.82128 ^c	8.83861 ^g	8.881 ^b
6	10.871199 ^c	10.87879 ^g	10.903 ^b

Table 1.1: Current best rigorous upper and lower bounds on the hypercubic lattice connective constant μ , together with estimates of actual values.

a) Conway and Guttmann (to be published), b) Alm (1992), c) Hara and Slade (1992b), d) Guttmann and Enting (1988), e) Guttmann (1987), f) Guttmann (1978), g) Guttmann (1981).

1.3 Generating functions

A common tool for understanding the behaviour of a sequence is its generating function. The generating function of the sequence $\{c_N\}$ is defined by

$$\chi(z) = \sum_{N=0}^{\infty} c_N z^N = \sum_{\omega} z^{|\omega|}. \quad (1.3.1)$$

The sum over ω is the sum over all self-avoiding walks, of arbitrary length $|\omega|$, which begin at the origin. The parameter z is known as the *activity*. Physically the activity occurs in the study of a canonical ensemble of polymers of variable length, and in this context is nonnegative. From a mathematical point of view, however, it will sometimes be useful to consider χ to be an analytic function of complex z .

Given two sites x and y , let $c_N(x, y)$ be the number of N -step self-avoiding walks ω with $\omega(0) = x$ and $\omega(N) = y$. The *two-point function* is the generating function for the sequence $c_N(x, y)$, i.e.,

$$G_z(x, y) = \sum_{N=0}^{\infty} c_N(x, y) z^N = \sum_{\omega: x \rightarrow y} z^{|\omega|}. \quad (1.3.2)$$

On the right side, the sum over ω is the sum over all self-avoiding walks, of arbitrary length, which begin at x and end at y . This is clearly translation invariant, so $G_z(x, y) = G_z(0, y - x)$. The two-point function is the self-avoiding walk analogue of the simple random walk Green function with

killing rate $1 - 2dz$:

$$C_z(x, y) = \sum_{N=0}^{\infty} p_N(x, y) (2dz)^N, \quad (1.3.3)$$

where $p_N(x, y)$ is the probability that an N -step simple random walk beginning at x ends at y .

The generating function for c_N can be written in terms of the two-point function as

$$\chi(z) = \sum_{x \in \mathbb{Z}^d} G_z(0, x). \quad (1.3.4)$$

In analogy with spin systems (see Section 2.3) we will refer to the generating function $\chi(z)$ as the *susceptibility*. The power series defining the susceptibility has radius of convergence

$$z_c \equiv \left[\lim_{N \rightarrow \infty} c_N^{1/N} \right]^{-1} = \frac{1}{\mu}, \quad (1.3.5)$$

and hence defines an analytic function in the *complex* parameter z if $|z| < z_c$. Since $c_N(0, x) \leq c_N$, the two-point function has radius of convergence at least z_c . It will be shown in Section 3.2 that in fact the radius of convergence is equal to z_c , for all $x \neq 0$. We will refer to z_c as the *critical point*, since it plays a role analogous to the critical point in statistical mechanical systems such as the Ising model or percolation.

It follows from (1.2.10) that

$$\chi(z) \geq \sum_{N=0}^{\infty} (\mu z)^N = \frac{1}{1 - \mu z} \quad (1.3.6)$$

and hence χ is “continuous” at the critical point, in the sense that $\chi(z) \rightarrow \infty$ as $z \nearrow z_c$. The manner of divergence of $\chi(z)$ at the critical point is related to the behaviour of the coefficients c_N for large N . To see this, we proceed as follows.

First we introduce the notation

$$f(x) \simeq g(x) \quad \text{as } x \rightarrow x_0 \quad (1.3.7)$$

to mean that there are positive constants C_1 and C_2 such that

$$C_1 g(x) \leq f(x) \leq C_2 g(x) \quad (1.3.8)$$

uniformly for x near its limiting value. Assuming that there is a γ such that

$$c_N \simeq \mu^N N^{\gamma-1} \quad \text{as } N \rightarrow \infty, \quad (1.3.9)$$

it can be concluded that

$$\chi(z) \simeq (z_c - z)^{-\gamma} \quad \text{as } z \nearrow z_c, \quad (1.3.10)$$

as follows. We write $z = \mu^{-1}e^{-t}$, so that $t \simeq z_c - z$. By the definition of $\chi(z)$,

$$\begin{aligned} \chi(z) &\simeq \sum_{N=1}^{\infty} N^{\gamma-1} e^{-tN} \simeq \int_1^{\infty} x^{\gamma-1} e^{-tx} dx \\ &= t^{-\gamma} \int_t^{\infty} y^{\gamma-1} e^{-y} dy \simeq t^{-\gamma}. \end{aligned}$$

In the above the sum can be replaced by the integral using Riemann sum approximations. The second integral converges as $t \searrow 0$, since by (1.2.10) $\gamma \geq 1$. Thus it is conjectured that

$$\chi(z) \sim A'(z_c - z)^{-\bar{\gamma}} \quad \text{as } z \nearrow z_c, \quad (1.3.11)$$

with $\bar{\gamma} = \gamma$.

As for the converse, it does not follow directly from (1.3.10) that (1.3.9) holds, without further assumptions. In general, the problem of extracting the large- n asymptotics of a sequence from the manner of divergence of its generating function is a Tauberian problem. An example of a Tauberian theorem providing a converse to the above argument will be given in Lemma 6.3.4.

Power law behaviour such as (1.3.10) is also observed for spin systems and percolation, and is characteristic of critical phenomena. It follows from (1.3.6) that $\bar{\gamma} \geq 1$, assuming that $\bar{\gamma}$ exists. In four dimensions, where it is believed that $c_N \sim A\mu^N(\log N)^{1/4}$, we expect similarly that $\chi(z) \sim A'(z_c - z)^{-1} |\log(z_c - z)|^{1/4}$.

The analogue $\chi_0(z)$ of $\chi(z)$ for simple random walk can be calculated explicitly:

$$\chi_0(z) = \sum_{N=0}^{\infty} (2dz)^N = \frac{1}{1-2dz}.$$

Thus the mean-field value of $\bar{\gamma}$ is 1, which not surprisingly is equal to the mean-field value for γ . The inequality $\bar{\gamma} \geq 1$ is an example of a *mean-field bound*. There is a sufficient condition for the opposite bound $\bar{\gamma} \leq 1$ known as the bubble condition, which is known to hold for $d \geq 5$ (and is believed not to hold for $d \leq 4$), and which will be discussed in detail in Section 1.5. There are many examples in critical phenomena of rigorous mean-field bounds, but, as mentioned in Section 1.1, no general proof is known of the mean-field bound $\nu \geq 1/2$.

We now turn our attention to the long distance behaviour of the two-point function. Below the critical point the two-point function decays exponentially. To see this, we note that $c_N(0, x) = 0$ for $N < \|x\|_{\infty}$, and hence

$$G_z(0, x) = \sum_{N=\|x\|_{\infty}}^{\infty} c_N(0, x) z^N \leq \sum_{N=\|x\|_{\infty}}^{\infty} c_N z^N. \quad (1.3.12)$$

Since $c_N^{1/N} \rightarrow \mu$ by (1.2.1), for any $\epsilon > 0$ there is a positive K_{ϵ} such that

$$c_N \leq K_{\epsilon}(\mu + \epsilon)^N \quad (1.3.13)$$

for all $N \geq 1$. Given a positive $z < z_c = \mu^{-1}$, we choose $\epsilon(z) > 0$ such that $\theta_z \equiv (\mu + \epsilon(z))z < 1$. Then substitution of (1.3.13) into (1.3.12) gives

$$G_z(0, x) \leq C_z \exp[-|\log \theta_z| \|x\|_{\infty}], \quad (1.3.14)$$

with $C_z = K_{\epsilon(z)}(1 - \theta_z)^{-1}$. This shows the desired exponential decay of the subcritical two-point function.

We define the *mass* $m(z)$ to be the rate of exponential decay of the two-point function along a coordinate axis:

$$m(z) = \liminf_{n \rightarrow \infty} \frac{-\log G_z(0, (n, 0, \dots, 0))}{n}. \quad (1.3.15)$$

In Theorem 4.1.3 it will be shown that in fact the $\liminf$ in the definition of m can be replaced by the limit. The *correlation length* $\xi(z)$, defined by $\xi(z) = m(z)^{-1}$, provides a characteristic length scale for the model.

The mass $m(z)$ is clearly not infinite for $0 < z < z_c$: considering only the shortest self-avoiding walk from 0 to $(n, 0, \dots, 0)$ gives

$$G_z(0, (n, 0, \dots, 0)) \geq z^n, \quad (1.3.16)$$

and hence $m(z) \leq -\log z$. By (1.3.14), $m(z) > 0$ for $z \in (0, z_c)$. By definition the mass is a nonincreasing function of positive z , and in Section 4.1 it will be shown that $m(z) \searrow 0$ as $z \nearrow z_c$. Given that the radius of convergence of $G_z(0, x)$ is z_c for all $x \neq 0$, it follows that $m(z) = -\infty$ for $z > z_c$. It has been proven that $m(z_c) = 0$ for $d \geq 5$; see Corollary 6.1.7. Although this is believed to be true in all dimensions, a negative mass at the critical point has not yet been ruled out rigorously in dimensions 2, 3 or 4.

Since the mass $m(z)$ approaches zero as $z \nearrow z_c$, it follows that the correlation length $\xi(z) = m(z)^{-1}$ diverges as $z \nearrow z_c$. It is believed that the manner of divergence of $\xi(z)$ is via a power law of the form

$$\xi(z) \sim \text{const.} (z_c - z)^{-\nu} \quad \text{as } z \nearrow z_c. \quad (1.3.17)$$

Formal scaling theory predicts that $\bar{\nu} = \nu$; this will be discussed in Section 2.1. The equality of these two critical exponents is part of a general belief that all length scales for the self-avoiding walk should be governed by the same critical exponent. The same belief generally applies to other statistical mechanical models as well.

Another correlation length, ξ_p , known as the *correlation length of order p* , is defined for each $p > 0$ by

$$\xi_p(z) = \left[\frac{\sum_{\omega} |\omega| (|\omega|)^p G_z(0, x)}{\sum_{\omega} z^{|\omega|}} \right]^{1/p} = \left[\frac{\sum_x |x|^p G_z(0, x)}{\sum_x G_z(0, x)} \right]^{1/p}. \quad (1.3.18)$$

By Hölder's inequality ξ_p is increasing in p . A formal argument similar to that showing $\bar{\nu} = \nu$ gives

$$\xi_p(z) \sim \text{const.} (z_c - z)^{-\nu_p} \text{ as } z \nearrow z_c,$$

with $\nu_p = \nu$ for all p .

For $p = 2$ there is no need to appeal to scaling theory to argue that $\nu_2 = \nu$. Instead we can argue as we did for the equality of γ and $\bar{\gamma}$, in the following way. We will assume that there exist exponents γ and ν such that $c_N \simeq \mu^N N^{\gamma-1}$ and $\langle |\omega|^2 \rangle \simeq N^{2\nu}$, and show that this implies that $\xi_2(z) \simeq (z_c - z)^{-\nu}$. Given the assumptions, we have

$$\begin{aligned} \sum_x |x|^2 G_z(0, x) &= \sum_N z^N \sum_{\omega: |\omega|=N} |\omega(N)|^2 \\ &\simeq \sum_N z^N N^{2\nu} c_N \simeq \sum_N z^N N^{2\nu+\gamma-1} \mu^N. \end{aligned} \quad (1.3.19)$$

Again writing $z = \mu^{-1} e^{-t}$, we obtain

$$\begin{aligned} \sum_x |x|^2 G_z(0, x) &\simeq \sum_N N^{2\nu+\gamma-1} e^{-tN} \\ &\simeq \int_1^\infty x^{2\nu+\gamma-1} e^{-tx} dx \simeq t^{-2\nu-\gamma}. \end{aligned} \quad (1.3.20)$$

This implies that

$$\xi_2(z)^2 \simeq \frac{t^{-2\nu-\gamma}}{\chi(z)} \simeq t^{-2\nu-\gamma+\bar{\gamma}}. \quad (1.3.21)$$

Using $\bar{\gamma} = \gamma$ it follows that $\xi_2(z) \simeq (z_c - z)^{-\nu}$, so $\nu_2 = \nu$.

1.4 Critical exponents

So far we have introduced the five critical exponents $\gamma, \bar{\gamma}, \nu, \bar{\nu}, \nu_p$. It was shown in Section 1.3 that if γ exists then $\gamma = \bar{\gamma}$. Heuristic arguments that $\nu = \bar{\nu} = \nu_p$ (for $0 < p < \infty$) will be given in the Section 2.1. The exponents were defined as follows:

$$c_N \sim A \mu^N N^{\gamma-1} \quad (1.4.1)$$

$$\chi(z) \sim A'(z_c - z)^{-\bar{\gamma}} \quad (1.4.2)$$

$$\langle |\omega(N)|^2 \rangle \sim DN^{2\nu} \quad (1.4.3)$$

$$\xi(z) \sim B(z_c - z)^{-\bar{\nu}} \quad (1.4.4)$$

$$\xi_p(z) \sim B_p(z_c - z)^{-\nu_p}. \quad (1.4.5)$$

We have written the above relations as if the various quantities involved are *asymptotically* given by power laws. This is consistent with the existing rigorous results, but some authors prefer a more conservative definition of the exponents. For example, one could require only that $c_N \simeq \mu^N N^{\gamma-1}$ [see (1.3.7)], with corresponding statements for the other exponents. A weaker definition, appearing sometimes in the literature, is to define the exponents by equations such as

$$\bar{\gamma} = - \lim_{z \nearrow z_c} \frac{\log \chi(z)}{\log(z_c - z)}, \quad (1.4.6)$$

but we will not need this definition. We shall take the optimistic view that the power law behaviour is asymptotic, although none of (1.4.1)–(1.4.5) has been proven in dimensions 2, 3, or 4 for any of these definitions of the exponents.

We will use the notation

$$f(x) \approx g(x) \quad (1.4.7)$$

in informal (nonrigorous) discussions to mean that $f(x)$ and $g(x)$ appear to have the same asymptotic behaviour in some sense which we will not attempt to specify.

In this section three additional critical exponents η , α_{sing} , and Δ_4 will be introduced. All these critical exponents are believed to be universal in the sense that they depend only on the dimension d of the lattice. In particular, the exponents are believed to be the same for the nearest-neighbour self-avoiding walk on $\mathbf{Z}^d$ as for a self-avoiding walk on $\mathbf{Z}^d$ in which steps can be within a fixed finite set $\Omega \subset \mathbf{Z}^d$ which is symmetric with respect to the symmetries of the lattice. Moreover the exponents ought even to be the same for a self-avoiding walk which can take unboundedly long steps, provided the weight of a step decays rapidly enough with its length (e.g.,

exponentially). This independence of the step set Ω is partially borne out in the rigorous results in high dimensions in Chapter 6.

We begin with the exponent η , which describes the conjectured long-distance behaviour of the two-point function at the critical point. Given that $m(z) \rightarrow 0$ as $z \nearrow z_c$, and given the belief that $m(z_c) = 0$ in all dimensions, it might be expected that the two-point function decays via a power law at the critical point. For simple random walk (with $d > 2$) the mass is certainly zero at the critical point, as it is well-known that the critical simple random walk two-point function $C_{1/2d}(0, x)$ decays like $|x|^{2-d}$ at large distances [see for example Lawler (1991)]. The conjectured large distance behaviour of the critical self-avoiding walk two-point function is

$$G_{z_c}(0, x) \sim \frac{C}{|x|^{d-2+\eta}} \quad \text{as } |x| \rightarrow \infty, \quad (1.4.8)$$

where C is a constant. This is believed to hold in all dimensions $d \geq 2$, including $d = 2$. Comparison with the simple random walk decay yields the mean field value 0 for η . Unfortunately it has not yet been proved rigorously that $G_{z_c}(0, x)$ is even finite for $d = 2, 3$ or 4, for any value of $x \neq 0$. For $d \geq 5$ somewhat weaker decay than (1.4.8) has been proved; see Theorem 6.1.6.

Assuming that (1.4.8) does provide the correct behaviour, it follows from the fact that the susceptibility is infinite at the critical point that $\eta \leq 2$. The value of η is believed to be determined from the values of γ and ν according to Fisher's scaling relation

$$\gamma = (2 - \eta)\nu. \quad (1.4.9)$$

The hypotheses leading to (1.4.9) will be discussed in Section 2.1. Inserting the conjectured values for γ and ν given in (1.1.11) and (1.1.12) into (1.4.9) gives the values for η appearing in Table 1.2. In contrast to γ and ν , the renormalization group predicts no logarithmic corrections to η in four dimensions. Logarithmic corrections are however expected in higher order terms in the asymptotic expansion of the critical two-point function in four dimensions.

One way to gain insight into the long distance behaviour of the critical two-point function is to examine the behaviour of its Fourier transform near the origin. In general, given a function $f(x)$ on the lattice whose absolute value is summable, we define its Fourier transform by

$$\hat{f}(k) = \sum_{x \in \mathbb{Z}^d} f(x) e^{ik \cdot x}, \quad k \in [-\pi, \pi]^d. \quad (1.4.10)$$

d	2	3	≥ 4
γ	$\frac{43}{32}$	1.162...	1
ν	$\frac{3}{4}$	0.59...	$\frac{1}{2}$
η	$\frac{5}{24}$	0.03...	0

Table 1.2: Conjectured values of γ, ν, η .

It is generally expected in critical phenomena that (1.4.8) is associated with behaviour of the form

$$\hat{G}_{z_c}(k) \sim \frac{C'}{k^{2-\eta}} \quad \text{as } k \rightarrow 0 \quad (1.4.11)$$

for some constant C' . [However (1.4.8) and (1.4.11) are not mathematically equivalent — an example of a function satisfying (1.4.11) but not (1.4.8) is given in the Notes at the end of the chapter.] Equation (1.4.11) has been established for $d \geq 5$ with $\eta = 0$ (see Theorem 6.1.6), but not yet for $d = 2, 3$ or 4. The conjectured values of η are all nonnegative. It is thus suggestive to conjecture the *infrared bound*

$$\hat{G}_z(k) \leq \frac{C}{k^2}, \quad (1.4.12)$$

with C independent of $k \in [-\pi, \pi]^d$ and $z \leq z_c$. For the nearest-neighbour Ising model and other reflection-positive spin systems the infrared bound is known rigorously to hold and was of considerable importance in the proof of mean-field behaviour of such models above four dimensions. For the self-avoiding walk it is still an open problem to prove the infrared bound in dimensions 2, 3 or 4, but in higher dimensions it has been proved (see Theorem 6.1.6). It is worth noting that the infrared bound is believed by some to be false for percolation and lattice animals below dimensions six and eight respectively (see Section 5.5 for more details about these models).

The exponent α_{sing} describes the behaviour of the number $c_N(0, x)$ of N -step self-avoiding walks which begin at the origin and end at x , as $N \rightarrow \infty$ with x fixed. For x equal to a nearest-neighbour e of the origin, $c_N(0, e)$ is closely related to the number of self-avoiding polygons. Self-avoiding polygons will be studied in detail in Section 3.2. It will be shown in Corollary 3.2.6 that the leading asymptotic behaviour of $c_N(0, x)$ as $N \rightarrow \infty$ is μ^N . As is the case for c_N , this leading behaviour is believed to

have a power law correction of the form

$$c_N(0, x) \sim B\mu^N N^{\alpha_{sing}-2}. \quad (1.4.13)$$

Here x is fixed and nonzero, and $N \rightarrow \infty$ through a sequence of values with the same parity as $\|x\|_1$. It is believed that α_{sing} is independent of x , and we formalize this conjecture for future reference as follows.

Conjecture 1.4.1 *For every pair of nonzero points x and y in $\mathbb{Z}^d$, there exist positive constants A_1 and A_2 and an integer N_0 (all depending on x and y) such that*

$$A_1 c_N(0, y) \leq c_N(0, x) \leq A_2 c_N(0, y) \quad \text{for all } N \geq N_0$$

if $\|x - y\|_1$ is even, and

$$A_1 c_{N+1}(0, y) \leq c_N(0, x) \leq A_2 c_{N+1}(0, y) \quad \text{for all } N \geq N_0$$

if $\|x - y\|_1$ is odd.

A special case of this conjecture is proven in Proposition 7.4.4. The value of B is also believed to be independent of x (as it is for simple random walk). For simple random walk the local central limit theorem states that the probability $p_N(0, x)$ that a simple random walk starting at 0 ends after N steps at x is given asymptotically by $\text{const.} N^{-d/2} \exp[-d|x|^2/2N] \sim \text{const.} N^{-d/2}$, as $N \rightarrow \infty$. Hence the mean-field value of $\alpha_{sing} - 2$ is $-d/2$. The value of α_{sing} is believed to be determined from the value of ν and the dimension d via the hyperscaling relation

$$\alpha_{sing} - 2 = -d\nu. \quad (1.4.14)$$

This hyperscaling relation will be discussed in Section 2.1. If (1.4.14) and the values given for ν in Table 1.2 are true, then it would follow that $\alpha_{sing} - 2 < -1$ in all dimensions and hence that the critical two-point function $G_{z_c}(0, x) = \sum_N c_N(0, x)\mu^{-N}$ is finite in all dimensions, including $d = 2$. This is in contrast to the situation for simple random walk, where in two dimensions the Green function is infinite at the critical point.

The strongest bounds on $c_N(0, x)$ are for high dimensions. It is proved in Theorem 6.1.3 that for d sufficiently large, or for $d > 4$ for a walk allowed to take long enough steps, that

$$c_N(0, x) \leq B\mu^N N^{-d/2} \quad (1.4.15)$$

for some constant B . Although this bound has not yet been extended to all $d \geq 5$ for the nearest-neighbour model, the weaker result that for all

$$a < -1 + d/2$$

$$\sup_x \sum_{N=0}^{\infty} N^a c_N(0, x)\mu^{-N} < \infty \quad (1.4.16)$$

has been proved for all $d \geq 5$; see Theorem 6.1.4. Either of (1.4.15) or (1.4.16) could be summarized by the inequality $\alpha_{sing} - 2 \leq -d/2$. For dimensions 2, 3 and 4, the best results are for x a nearest neighbour of the origin. These results are described in Section 8.1 and can be summarized by the inequalities

$$\alpha_{sing} \leq \frac{5}{2} \quad (d = 2) \quad (1.4.17)$$

$$\alpha_{sing} \leq 2 \quad (d = 3) \quad (1.4.18)$$

$$\alpha_{sing} < 2 \quad (d \geq 4). \quad (1.4.19)$$

Finally, we introduce the critical exponent Δ_4 . Let $c_{N_1, N_2}(x)$ denote the number of pairs of self-avoiding walks of lengths N_1 and N_2 and respective starting points 0 and x which intersect each other, and let

$$c_{N_1, N_2} = \sum_x c_{N_1, N_2}(x). \quad (1.4.20)$$

This quantity occurs in the study of interacting polymer chains. The asymptotic behaviour of c_{N_1, N_2} is believed to be given by

$$c_{N_1, N_2} \sim \text{const.} \mu^{N_1+N_2} N_1^{2\Delta_4+\gamma-2} f(N_1/N_2) \quad \text{as } N_1, N_2 \rightarrow \infty \quad (1.4.21)$$

for some critical exponent Δ_4 and universal scaling function f . The quantity

$$g(z) = \xi(z)^{-d} \chi(z)^{-2} \sum_{N_1, N_2=0}^{\infty} c_{N_1, N_2} z^{N_1+N_2} \quad (1.4.22)$$

represents a kind of average intersection probability. In quantum field theory, an analogue of $g(z)$ is referred to as the renormalized coupling constant. An informal calculation in which (1.4.21) is substituted into (1.4.22) leads to

$$g(z) \sim \text{const.} (z_c - z)^{d\nu-2\Delta_4+\gamma} \quad \text{as } z \nearrow z_c. \quad (1.4.23)$$

For simple random walk it is known that the analogue of $g(z)$ satisfies (1.4.23), with $d/2 - 2\Delta_4 + 1 = 0$ for $d = 2, 3, 4$ (with a logarithmic correction in four dimensions), and $\Delta_4 = 3/2$ for $d \geq 5$; see Section 10.3. Similar behaviour is believed to hold for the self-avoiding walk. In particular, for the self-avoiding walk it is believed that in dimensions 2, 3 and 4 the hyperscaling relation

$$d\nu - 2\Delta_4 + \gamma = 0 \quad (1.4.24)$$

is satisfied. Heuristic arguments in support of this hyperscaling relation will be given in Section 2.1. It has been proved that $\Delta_4 = 3/2$ for the self-avoiding walk in dimensions $d \geq 6$ (see Theorem 1.5.5 and the Remark following its statement); it is believed that $\Delta_4 = 3/2$ for all $d > 4$.

Elementary bounds on Δ_4 can be obtained as follows. Consider all pairs of N -step self-avoiding walks $\omega^{(1)}$ and $\omega^{(2)}$ which intersect somewhere, with $\omega^{(1)}$ beginning at the origin and $\omega^{(2)}$ beginning anywhere. There are $c_{N,N}$ such pairs. Since there are $N+1$ possible sites on each of $\omega^{(1)}$ and $\omega^{(2)}$ where an intersection can occur, $c_{N,N} \leq (N+1)^2 c_N^2$. On the other hand if we count only those pairs for which $\omega^{(2)}(0) = \omega^{(1)}(j)$ for some $j = 0, \dots, n$, we obtain $c_{N,N} \geq (N+1)c_N^2$. Together these bounds give

$$\frac{\gamma+1}{2} \leq \Delta_4 \leq \frac{\gamma+2}{2}. \quad (1.4.25)$$

This can be rewritten as

$$1 \leq 2\Delta_4 - \gamma \leq 2. \quad (1.4.26)$$

The upper bound implies that the hyperscaling relation (1.4.24) fails if $d\nu > 2$. Since it is known that $\nu = 1/2$ for $d \geq 5$ (see Section 6.1), this implies failure of hyperscaling for $d > 4$.

1.5 The bubble condition

The lower bound on the susceptibility (1.3.6) can be rewritten in terms of $z_c = \mu^{-1}$ as

$$\chi(z) \geq \frac{z_c}{z_c - z} \quad (1.5.1)$$

for $0 \leq z < z_c$. The bubble condition is a sufficient condition for the complementary bound

$$\chi(z) \leq \frac{C}{z_c - z} \quad (1.5.2)$$

for some constant C and for $0 \leq z < z_c$. Thus the bubble condition implies that $\bar{\gamma} = 1$ in the sense that

$$\chi(z) \simeq (z_c - z)^{-1} \quad \text{as } z \nearrow z_c. \quad (1.5.3)$$

The bubble condition was proven to hold in five or more dimensions in Hara and Slade (1992b) (see Section 6.1), and is believed not to hold for $d \leq 4$.

To state the bubble condition we first introduce the *bubble diagram*

$$B(z) = \sum_{x \in \mathbb{Z}^d} G_z(0, x)^2. \quad (1.5.4)$$

1.5. THE BUBBLE CONDITION

The name “bubble diagram” comes from a Feynman diagram notation in which the two-point function or *propagator* evaluated at sites x and y is denoted by a line terminating at x and y . In this notation

$$B(z) = \sum_x 0 \circ x = \circ$$

where in the diagram on the right it is implicit that one vertex is fixed at the origin and the other is summed over the lattice. The bubble diagram can be rewritten in terms of the Fourier transform of the two-point function, using (1.5.4) and the Parseval relation, as

$$B(z) = \|G_z(0, \cdot)\|_2^2 = \|\hat{G}_z\|_2^2 = \int_{[-\pi, \pi]^d} \hat{G}_z(k)^2 \frac{d^d k}{(2\pi)^d}. \quad (1.5.5)$$

Definition 1.5.1 *The bubble condition states that the bubble diagram is finite at the critical point, i.e.*

$$B(z_c) < \infty.$$

In view of the definition of η in (1.4.8) or (1.4.11), it follows from (1.5.5) that the bubble condition is satisfied provided $\eta > (4-d)/2$. Hence the bubble condition for $d > 4$ is implied by the infrared bound $\eta \geq 0$. If the values for η given in Table 1.2 are correct, then the bubble condition will not hold in dimensions 2, 3 or 4, with the divergence of the bubble diagram being only logarithmic in four dimensions.

The next lemma provides the principal step in proving that the bubble condition implies (1.5.2) and hence implies (1.5.3).

Lemma 1.5.2 *For any $z \in [0, z_c)$, the derivative of the susceptibility satisfies*

$$\frac{\chi(z)^2}{B(z)} - \chi(z) \leq z\chi'(z) \leq \chi(z)^2 - \chi(z). \quad (1.5.6)$$

Proof. Below the critical point the derivative of χ can be obtained by term by term differentiation:

$$z\chi'(z) = \sum_{\omega} |\omega| z^{|\omega|} = \sum_{\omega} (|\omega| + 1) z^{|\omega|} - \chi(z), \quad (1.5.7)$$

where the sums are over self-avoiding walks of arbitrary length which begin at the origin. The summation on the right side can be written

$$\sum_y \sum_{\omega: 0 \rightarrow y} \sum_x I[\omega(j) = x \text{ for some } j] z^{|\omega|}$$

$$\begin{aligned}
&= \sum_{x,y} \sum_{\substack{\omega^{(1)}: 0 \rightarrow x \\ \omega^{(2)}: x \rightarrow y}} z^{|\omega^{(1)}|+|\omega^{(2)}|} I[\omega^{(1)} \cap \omega^{(2)} = \{x\}] \\
&\equiv Q(z),
\end{aligned} \tag{1.5.8}$$

where I denotes the indicator function and the last summation is over self-avoiding walks $\omega^{(1)}$ and $\omega^{(2)}$ of arbitrary length and having the prescribed endpoints. Then

$$z\chi'(z) = Q(z) - \chi(z). \tag{1.5.9}$$

The upper bound in (1.5.6) then follows since the indicator function in the middle member of (1.5.8) is bounded above by one.

The first step toward obtaining the lower bound is to use the inclusion-exclusion relation in the form

$$I[\omega^{(1)} \cap \omega^{(2)} = \{x\}] = 1 - I[\omega^{(1)} \cap \omega^{(2)} \neq \{x\}].$$

This gives

$$Q(z) = \chi(z)^2 - \sum_{x,y} \sum_{\substack{\omega^{(1)}: 0 \rightarrow x \\ \omega^{(2)}: x \rightarrow y}} z^{|\omega^{(1)}|+|\omega^{(2)}|} I[\omega^{(1)} \cap \omega^{(2)} \neq \{x\}]. \tag{1.5.10}$$

In the last term on the right side of (1.5.10), let $w = \omega^{(2)}(l)$ be the site of the last intersection of $\omega^{(2)}$ with $\omega^{(1)}$, where time is measured along $\omega^{(2)}$ beginning at its starting point x . Then the portion of $\omega^{(2)}$ corresponding to times greater than l must avoid all of $\omega^{(1)}$. Relaxing the restrictions that this portion of $\omega^{(2)}$ avoid both the remainder of $\omega^{(2)}$ and the part of $\omega^{(1)}$ linking w to x gives the upper bound

$$\sum_{x,y} \sum_{\substack{\omega^{(1)}: 0 \rightarrow x \\ \omega^{(2)}: x \rightarrow y}} z^{|\omega^{(1)}|+|\omega^{(2)}|} I[\omega^{(1)} \cap \omega^{(2)} \neq \{x\}] \leq Q(z)[B(z) - 1]. \tag{1.5.11}$$

Here the factor $B(z) - 1$ arises from the two paths joining w and x . The upper bound involves $B(z) - 1$ rather than $B(z)$ since there will be no contribution here from the $x = 0$ term in (1.5.4). This type of distinction will be crucial in similar bounds on the lace expansion used in Chapter 6.

Combining (1.5.10) and (1.5.11) gives

$$Q(z) \geq \chi(z)^2 - Q(z)[B(z) - 1]. \tag{1.5.12}$$

This inequality is illustrated in Figure 1.2. Solving for $Q(z)$ gives

$$Q(z) \geq \frac{\chi(z)^2}{B(z)}.$$

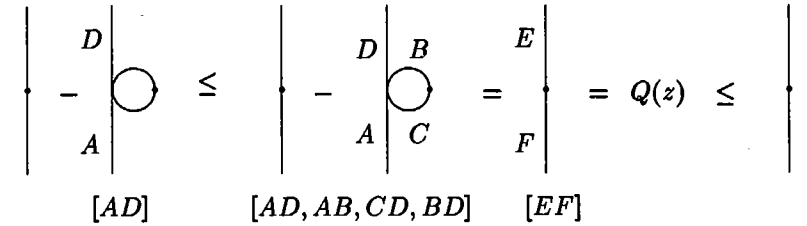


Figure 1.2: A diagrammatic representation of the inequality $\chi(z)^2 - Q(z)[B(z) - 1] \leq Q(z) \leq \chi(z)^2$ occurring in the proof of Lemma 1.5.2. The list of pairs of lines indicates interactions between the propagators, in the sense that the corresponding walks must avoid each other.

Combining this inequality with (1.5.9) completes the proof of the lemma. $\square$

The quantity $\chi(z)^{-2}Q(z)$ can be interpreted as the probability that two self-avoiding walks of arbitrary length, which start at the origin, do not intersect. Lemma 1.5.2 can be restated as saying that this probability lies in the interval $[B(z)^{-1}, 1]$, and hence remains strictly positive at the critical point if the bubble condition is satisfied.

In the next theorem it is shown that the lower bound of Lemma 1.5.2 implies that if the bubble condition is satisfied, then (1.5.3) holds.

Theorem 1.5.3 *If the bubble condition is satisfied, and hence in particular if the infrared bound holds and $d > 4$, then there is a positive function $\epsilon(z)$ with $\lim_{z \nearrow z_c} \epsilon(z) = 0$ such that for z less than but near z_c*

$$\frac{z_c}{z_c - z} \leq \chi(z) \leq \frac{z_c[B(z_c) + \epsilon(z)]}{z_c - z}.$$

Hence if in fact there is a constant A such that $\chi(z) \sim Az_c(z_c - z)^{-1}$, then $1 \leq A \leq B(z_c)$.

Proof. The lower bound in the statement of the theorem is just (1.5.1). For the upper bound, let $z_1 \in (0, z_c)$. It follows from the lower bound in (1.5.6) that for $z \in [z_1, z_c)$

$$\begin{aligned}
z \left(-\frac{d\chi^{-1}}{dz} \right) &\geq \frac{1}{B(z)} - \frac{1}{\chi(z)} \\
&\geq \frac{1}{B(z_c)} - \frac{1}{\chi(z_1)}.
\end{aligned} \tag{1.5.13}$$

We bound the factor of z on the left side by z_c and then integrate from z_1 to z_c . Using the fact that $\chi(z_c)^{-1} = 0$ by (1.5.1), this gives

$$z_c \chi(z_1)^{-1} \geq [B(z_c)^{-1} - \chi(z_1)^{-1}](z_c - z_1). \quad (1.5.14)$$

Rewriting gives

$$\chi(z_1) \leq \frac{B(z_c)}{1 - B(z_c)\chi(z_1)^{-1}} \frac{z_c}{z_c - z_1}. \quad (1.5.15)$$

This gives the desired upper bound on the susceptibility, since by (1.5.1) the inverse susceptibility on the right side can be made arbitrarily small by taking z_1 sufficiently close to z_c . $\square$

Although the bubble condition is expected not to hold in four dimensions, it is nevertheless possible to draw some conclusions from the lower bound of Lemma 1.5.2 if we assume the infrared bound (1.4.12). While not sharp compared to the expected behaviour

$$\chi(z) \sim \frac{A}{z_c - z} |\log(z_c - z)|^{1/4},$$

the upper bound that we obtain on χ shows that the deviation from mean-field behaviour is at worst logarithmic in four dimensions, if the infrared bound is satisfied.

Theorem 1.5.4 *Let $d = 4$. If the infrared bound (1.4.12) is satisfied then for z less than but near z_c ,*

$$\frac{z_c}{z_c - z} \leq \chi(z) \leq C \frac{|\log(z_c - z)|}{z_c - z}$$

for some constant C which does not depend on z .

Proof. The lower bound in the statement of the theorem is just (1.5.1), which holds in all dimensions. It remains to prove the upper bound. In the following, C represents a constant whose value may change from one occurrence to another.

Let $0 < z < z_c$. Since

$$\chi(z) = \hat{G}_z(0) \geq |\hat{G}_z(k)|$$

for all k , it follows from the infrared bound that

$$|\hat{G}_z(k)| \leq \frac{2}{|\hat{G}_z(k)|^{-1} + \chi(z)^{-1}} \leq \frac{2C}{k^2 + C\chi(z)^{-1}}. \quad (1.5.16)$$

Using the fact that

$$B(z) = \int_{[-\pi, \pi]^4} \hat{G}_z(k)^2 \frac{d^4 k}{(2\pi)^4},$$

a routine calculation using (1.5.16) gives the bound

$$B(z) \leq C[1 + \log \chi(z)]. \quad (1.5.17)$$

By (1.5.6), (1.5.1) and (1.5.17), for z sufficiently close to z_c we have

$$\begin{aligned} -z \frac{d\chi^{-1}}{dz} &\geq \frac{1}{B(z)} - \frac{1}{\chi(z)} \\ &\geq \frac{1}{2B(z)} \\ &\geq \frac{C}{1 + \log \chi(z)} \end{aligned}$$

and therefore

$$-[1 + \log \chi(z)] \frac{d\chi^{-1}}{dz} \geq C. \quad (1.5.18)$$

The left side of (1.5.18) is the derivative of $-\chi(z)^{-1}[2 + \log \chi(z)]$. Hence for z close to z_c integration of (1.5.18) over the interval (z, z_c) gives

$$\chi(z)^{-1}[2 + \log \chi(z)] \geq C(z_c - z),$$

where we used (1.5.1) to see that the contribution from the upper limit of integration on the left side is zero. Decreasing C slightly we obtain

$$\frac{1}{C(z_c - z)} \geq \chi(z)[\log \chi(z)]^{-1}. \quad (1.5.19)$$

Taking logarithms, and taking z sufficiently close to z_c , gives

$$C|\log(z_c - z)| \geq \log \chi(z) - \log \log \chi(z) \geq \frac{1}{2} \log \chi(z). \quad (1.5.20)$$

Inserting the lower bound for $[\log \chi(z)]^{-1}$ given by (1.5.20) into (1.5.19) gives

$$C(z_c - z)^{-1} \geq \chi(z)|\log(z_c - z)|^{-1}.$$

This gives the upper bound on χ in the statement of the theorem. $\square$

Finally we turn to a connection between the bubble diagram and the critical exponent Δ_4 for the renormalized coupling constant $g(z)$, which was defined in (1.4.22) by

$$g(z) = \xi(z)^{-d} \chi(z)^{-2} \sum_{N_1, N_2=0}^{\infty} c_{N_1, N_2} z^{N_1 + N_2}. \quad (1.5.21)$$

Here c_{N_1, N_2} is the sum over sites x of the number of intersecting pairs of self-avoiding walks of length N_1 and N_2 starting at 0 and x respectively. The critical behaviour of $g(z)$ is believed to be of the form $(z_c - z)^{d\nu - 2\Delta_4 + \gamma}$.

The next theorem gives sufficient conditions for Δ_4 to take its mean-field value $3/2$. The theorem is most efficiently stated in terms of the *repulsive* bubble diagram $R(z) < B(z)$, which is defined by taking only those contributions to the bubble from pairs of walks which are mutually avoiding apart from their common endpoints:

$$R(z) = \sum_{x \in \mathbb{Z}^d} \sum_{\substack{\omega^{(1)}: 0 \rightarrow x \\ \omega^{(2)}: 0 \rightarrow x}} z^{|\omega^{(1)}| + |\omega^{(2)}|} I[\omega^{(1)} \cap \omega^{(2)} = \{0, x\}]. \quad (1.5.22)$$

Theorem 1.5.5 *If $B(z_c) < \infty$ and in addition $R(z_c) - 1 < 1/4$, then $g(z) \simeq \xi(z)^{-d}(z_c - z)^{-2}$. If also $\xi(z) \simeq (z_c - z)^{-\nu}$, then $\Delta_4 = 3/2$ in the sense that*

$$g(z) \simeq (z_c - z)^{d\nu - 3 + \gamma} = (z_c - z)^{d\nu - 2}$$

(assuming that the exponent γ for c_n is equal to the exponent for the susceptibility).

Remark. The best current bound on $R(z_c) - 1$ in five dimensions is $0.434636 > 0.25$ [Hara and Slade (1991b)]. For $d = 6$ the same reference reports $B(z_c) - 1 \leq 0.25974$. However the repulsive bubble in six dimensions satisfies $R(z_c) - 1 \leq 0.2343 < 0.25$, and is smaller still in more than six dimensions [Hara and Slade (unpublished)]. Together with Theorem 1.5.5 and the result of Hara and Slade (1991a) that for $d \geq 5$ the correlation length exhibits the mean-field behaviour $\xi(z) \sim \text{const.}(z_c - z)^{-1/2}$ (and that the exponent for c_n is $\gamma = 1$), this implies that

$$g(z) \simeq (z_c - z)^{(d-4)/2} \quad (1.5.23)$$

for $d \geq 6$. Although the same conclusion cannot yet be made for $d = 5$, it will be shown in Chapter 6 (see Theorem 6.2.5 and the remark preceding it, and Theorem 6.1.5) that for a sufficiently spread-out self-avoiding walk in more than four dimensions, $B(z_c) - 1 < 1/4$ and $\xi(z) \sim \text{const.}(z_c - z)^{-1/2}$, and hence $g(z) \simeq (z_c - z)^{(d-4)/2}$.

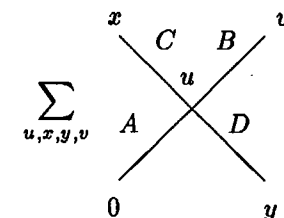
Proof of Theorem 1.5.5. By Theorem 1.5.3, the bubble condition implies that $\chi(z) \simeq (z_c - z)^{-1}$. Hence to prove the theorem it suffices to show that

$$\sum_{N_1, N_2=0}^{\infty} c_{N_1, N_2} z^{N_1 + N_2} \simeq (z_c - z)^{-4}. \quad (1.5.24)$$

The left side is equal to

$$\sum_{x, y, v} \sum_{\substack{\omega: 0 \rightarrow v \\ \rho: x \rightarrow y}} z^{|\omega| + |\rho|} I[\omega \cap \rho \neq \emptyset]. \quad (1.5.25)$$

In a nonzero contribution to this sum, let u be the first site along ω where ω and ρ intersect. Then the portion of ω before u avoids ρ as well as the latter part of ω , while the latter part of ω avoids only the former part of ω and may intersect ρ . This gives the following diagrammatic interpretation of the left side of (1.5.24) (in which the list of pairs indicates mutually avoiding walks):



$$[AB, CD, AC, AD] \quad (1.5.26)$$

Neglecting all mutual avoidance between the four lines of the diagram gives the upper bound $\chi^4 \leq \text{const.}(z_c - z)^{-4}$ for the left side of (1.5.24).

For a lower bound on (1.5.26) we apply inclusion-exclusion, as follows. The indicator function for the event that the various mutual avoidances shown in (1.5.26) occur can be written as one minus the event that at least one of the required mutual avoidances is violated. This leads to the lower bound

$$\begin{aligned} & \sum_{u, v, x, y} \sum_{\substack{\omega^{(1)}: 0 \rightarrow u \\ \omega^{(2)}: u \rightarrow v}} \sum_{\substack{\rho^{(1)}: x \rightarrow u \\ \rho^{(2)}: u \rightarrow y}} z^{|\omega^{(1)}| + |\omega^{(2)}| + |\rho^{(1)}| + |\rho^{(2)}|} \\ & \times \left\{ 1 - I[\omega^{(1)} \cap \omega^{(2)} \neq \{u\}] - I[\omega^{(1)} \cap \rho^{(1)} \neq \{u\}] \right. \\ & \left. - I[\omega^{(1)} \cap \rho^{(2)} \neq \{u\}] - I[\rho^{(1)} \cap \rho^{(2)} \neq \{u\}] \right\}. \quad (1.5.27) \end{aligned}$$

This bound is equal to

$$\chi^4 - 4\chi^2 \sum_{u, x} \sum_{\substack{\gamma^{(1)}: 0 \rightarrow u \\ \gamma^{(2)}: u \rightarrow x}} z^{|\gamma^{(1)}| + |\gamma^{(2)}|} I[\gamma^{(1)} \cap \gamma^{(2)} \neq \{u\}]. \quad (1.5.28)$$

We now argue as in (1.5.11), but this time we let w be the site of the *first* intersection (measured along $\gamma^{(2)}$) of $\gamma^{(2)}$ with $\gamma^{(1)}$. This gives the lower bound

$$\chi^4 - 4\chi^4[R(z) - 1] \geq \text{const.}\chi^4 \quad (1.5.29)$$

for (1.5.26), assuming that $R(z_c) - 1 < 1/4$. $\square$

1.6 Notes

Section 1.1. Existence of the connective constant $\mu = \lim_{N \rightarrow \infty} c_N^{1/N}$ was first proven in Hammersley and Morton (1954); this paper essentially marks the beginning of rigorous results for the self-avoiding walk. The (nonrigorous) derivation of $\mu = \sqrt{2 + \sqrt{2}}$ for the honeycomb lattice is due to Nienhuis (1982); see also Nienhuis (1984) and Nienhuis (1987). For high dimensions, it was shown in Kesten (1964) that $\mu = 2d - 1 - (2d)^{-1} + O(d^{-2})$, and this has recently been improved to $\mu = 2d - 1 - (2d)^{-1} - 3(2d)^{-2} + O(d^{-3})$ using the lace expansion [Hara and Slade (unpublished)].

The conjectured values for γ and ν in two dimensions arise from an exact solution which is described in the articles by Nienhuis cited above. An alternate approach, based on conformal invariance, is discussed in Duplantier (1989), Duplantier (1990), and references therein. A rigorous argument leading to these two-dimensional critical exponents remains an open problem of major importance, and a solution would likely have far-reaching implications. For $d = 3$, field theoretic computations of the critical exponents are given in Le Guillou and Zinn-Justin (1989). Monte Carlo computations of the exponents are given for example in Madras and Sokal (1988), and numerical computations using extrapolation of exact enumerations are given in Guttmann and Wang (1991). The logarithmic corrections in four dimensions are obtained in Larkin and Khmel'Nitskii (1969), Wegner and Riedel (1973) and Brezin, Le Guillou and Zinn-Justin (1973). For recent progress on rigorous results in four dimensions, see Brydges, Evans and Imbrie (1992) and Arnaudon, Iagolnitzer and Magnen (1991). Existence of critical exponents for $d \geq 5$ is proven in Hara and Slade (1992a, 1992b).

A necessary and sufficient condition for a bound of the form $c_N \leq \text{const.}\mu^N N^H$ for some finite H , i.e. for the finiteness of the critical exponent γ , is given in Hammersley (1991, 1992). We note the presence of a minor error in Hammersley (1991): the right side of (30) does not follow from the inequality that precedes it. This is easy to fix, however, as follows. In Hammersley's notation, the bound $f(m) \leq Gm^H \mu^m$ implies that $f(m, r) \leq \sum G^r n_1^H n_2^H \dots n_r^H \mu^m$, where the sum is over all $n_1, \dots, n_r \geq 1$ that sum to m . By the arithmetic-geometric inequality we have $n_1 n_2 \dots n_r \leq (m/r)^r$,

which implies $f(m, r) \leq \binom{m-1}{r-1} G^r (m/r)^r \mu^m$. This gives us (30) with $Hu \log H$ replaced by $-Hu \log u$, and Hammersley's Equation (3) follows.

A rigorous understanding of the self-avoiding walk on finitely ramified fractals has recently emerged; see Hattori (1992) for a review.

For the Ising model (and also for φ^4 field theory), the following references prove results concerning mean-field behaviour above four dimensions: Sokal (1979), Aizenman (1982), Fröhlich (1982), Aizenman and Fernández (1986), Fernández, Fröhlich and Sokal (1992).

Section 1.2. The bound $(c_N/c_1)^{1/(N-1)}$ (for all $N \geq 2$) is attributed to Alm in Ahlberg and Janson (1980). The latter reference obtains an improvement when $c_N/c_{N-1} > c_1 - 2$: they show that μ is bounded above by the unique positive root of the polynomial

$$c_1 x^{N-1} = [c_N - (c_1 - 2)c_{N-1}]x + (c_1 - 2)[(c_1 - 1)c_{N-1} - c_N] \quad (1.6.1)$$

(for all $N \geq 3$). Currently the best upper bounds available are due to Alm (1992).

A method for obtaining lower bounds on μ using bridges was given in Guttmann (1983). The current best lower bound in two dimensions, due to Conway and Guttmann (to be published), also uses bridges. For $d \geq 3$, the best lower bounds are due to Hara and Slade (1992b), who use a different approach involving loop erasure.

The numerical estimates for μ cited in Table 1.1 are from exact enumeration data.

Section 1.3. Exponential decay of the subcritical two-point function was proven in Fisher (1966), as part of a study of the form of the distribution of $c_N(0, x)$.

Section 1.4. We make no attempt here to refer to the original literature on critical exponents; the ideas in this section are part of the standard physics picture of critical phenomena.

The infrared bound was proven for reflection-positive spin systems in all dimensions in Fröhlich, Simon and Spencer (1976). For branched polymers and for percolation, there are arguments that the infrared bound does not hold below eight and six dimensions respectively; see Bovier, Fröhlich and Glauber (1986) and Adler (1984) respectively.

For $d \geq 5$ it has been proven that $\tilde{G}_{z_c}(k) \sim \text{const.}k^{-2}$ as $k \rightarrow 0$, but although it is believed that $G_{z_c}(x)$ is asymptotic to a multiple of $|x|^{2-d}$, this has not yet been proven (see Theorem 6.1.6 for a weaker result). It is thus of interest to know under what conditions behaviour of the form

$k^{-2+\eta}$ for a Fourier transform $\hat{g}(k)$ implies behaviour of the form $|x|^{2-d-\eta}$ for $g(x)$. For the case $\eta = 0$, the following sufficient condition was pointed out to us by S. Kotani (private communication); we omit the proof.

Theorem 1.6.1 *Let $d \geq 3$, and let $\mathbf{T}^d \equiv (\mathbf{R}/2\pi\mathbf{Z})^d$. Let $\hat{g}$ be a function in $C^{d-2}(\mathbf{T}^d \setminus \{0\})$, let $\hat{h}(k) = k^2 \hat{g}(k)$, and for $x \in \mathbf{Z}^d$ let $g(x) = (2\pi)^{-d} \int_{\mathbf{T}^d} \hat{g}(k) e^{-ik \cdot x} d^d k$. Suppose that there is a neighbourhood $U \subset \mathbf{T}^d$ of 0 such that*

$$\hat{h} \in \begin{cases} C^{d-1}(U) & \text{if } d = 3, 4 \\ C^{d-2}(U) & \text{if } d \geq 5. \end{cases}$$

Then as $|x| \rightarrow \infty$,

$$g(x) \sim \hat{h}(0) \frac{\Gamma(d/2)}{2(d-2)\pi^{d/2}} |x|^{-(d-2)}.$$

The following shows that in general the hypothesis of existence of $d-2$ derivatives for $\hat{h}$ cannot be relaxed: we give an example² of a function $\hat{g}$ on $\mathbf{T}^d$, for $d \geq 3$, which is asymptotic to a multiple of k^{-2} as $k \rightarrow 0$, with $\hat{h}(k) = k^2 \hat{g}(k)$ having $d-3$ but not $d-2$ derivatives in a neighbourhood of $k = 0$, but for which $g(x)$ is not bounded above by a multiple of $|x|^{2-d}$ for large x .

Example 1.6.2 Let $d \geq 3$, and let $C(x)$ be the critical simple random walk two-point function (or in other words the Green function) studied in Appendix A. Then $C(x)$ is asymptotic to a multiple of $|x|^{2-d}$ for large x [see, e.g., Lawler (1991)]. Also, $\hat{C}(k)^{-1} = 1 - d^{-1} \sum_{\mu=1}^d \cos k_\mu$ is asymptotic to $(2d)^{-1} k^2$ as $k \rightarrow 0$. Fix q such that $d-3 < q < d-2$, and for $-\pi \leq t \leq \pi$ define

$$\hat{f}(t) = \sum_{m=-\infty}^{\infty} 2^{-q|m|} \exp[it(\operatorname{sgn} m)2^{|m|}], \quad (1.6.2)$$

where $\operatorname{sgn} m = +1$ if $m > 0$; $= 0$ if $m = 0$; $= -1$ if $m < 0$. For $k \in [-\pi, \pi]^d$, let

$$\hat{F}(k) = \epsilon \prod_{\mu=1}^d \hat{f}(k_\mu) \quad (1.6.3)$$

where ϵ is chosen small enough that $1 + \hat{C}(k)^{-1} \hat{F}(k)$ is strictly positive uniformly in all $k \in [-\pi, \pi]^d$. (This is possible since $\hat{C}(k)^{-1}$ and the product in (1.6.3) are both bounded uniformly in k .) Observe that $\hat{F} \in C^s(\mathbf{T}^d)$ for $s < q$, but that for $s > q$, $\partial_\mu^s \hat{F}(0)$ does not exist. Now let

$$\hat{g}(k) = \hat{C}(k) + \hat{F}(k) = \hat{C}(k)[1 + \hat{C}(k)^{-1} \hat{F}(k)] \quad (1.6.4)$$

²The example was arrived at in conversation with T. Hara.

and

$$\hat{h}(k) = k^2 \hat{g}(k). \quad (1.6.5)$$

Then $\hat{g}(k)$ is asymptotic to $(2d)k^{-2}$ as $k \rightarrow 0$, and $\hat{h}(k) \in C^{d-3}(\mathbf{T}^d)$. However $g(x)$ is not bounded above by a multiple of $|x|^{2-d}$ for large x , because $F(x) = \epsilon|x|^{-q}$ for x having one component of the form $\pm 2^{|m|}$ (for any integer m) and all other components zero.

See Appendix A of Sokal (1982) for a discussion of some related issues.

Section 1.5. For reflection positive spin systems the infrared bound was proven in Fröhlich, Simon and Spencer (1976). As a consequence the bubble diagram for such systems is finite at the critical point above four dimensions, and diverges logarithmically in four dimensions. This was used to prove mean-field behaviour for spin systems for dimensions greater than four in Aizenman (1982) and Fröhlich (1982). In Bovier, Felder and Fröhlich (1984) Theorem 1.5.3 was proved, although at that time for the self-avoiding walk neither the infrared bound nor the bubble condition were known to hold in any dimension. In the same paper it was observed that if the infrared bound holds in four dimensions then the deviation from mean-field behaviour for the susceptibility is at most logarithmic. Our proof of Theorem 1.5.4 yields this conclusion in a slightly stronger form, following the methods used for spin systems in Aizenman and Graham (1983). Results analogous to Theorem 1.5.5 were obtained for spin systems in Aizenman (1982) and Fröhlich (1982). The proof that $\Delta_4 = 3/2$ for $d \geq 6$ is new, and is due to Hara and Slade (unpublished).

For percolation and branched polymers (lattice trees and lattice animals) the role of the bubble diagram is played by the triangle and the square diagram respectively; see Section 5.5. For percolation see Aizenman and Newman (1984), Nguyen (1987), Barsky and Aizenman (1991), Hara and Slade (1990a) and Nguyen and Yang (1991). For lattice trees and lattice animals see Bovier, Fröhlich and Glaes (1986), Tasaki and Hara (1987) and Hara and Slade (1990b).

Chapter 2

Scaling, polymers and spins

2.1 Scaling theory

This chapter is concerned with some of the nonrigorous work on the self-avoiding walk: the scaling theory which leads to the scaling relations stated in Section 1.4, the connection with polymers and the derivation of the Flory values for the critical exponent ν , and finally, the interpretation of the self-avoiding walk as a “zero-component” ferromagnet.

We begin in this section with a discussion of scaling theory, giving heuristic derivations of Fisher’s scaling relation (1.4.9) and of the hyperscaling relations (1.4.14) and (1.4.24). There are a variety of approaches which can be used to derive these relations, and here we content ourselves with giving a representative sample of the types of arguments which are frequently used. Although the sort of arguments we will describe are part of the standard lore of theoretical physics (applicable to a wide variety of models), from a mathematical point of view they may appear to be on rather shaky ground. There will be no rigorous results in this section, and we will make frequent use of the symbol $\approx$, which implies a leap of faith.

Our starting point will be an assumption about the behaviour of the two-point function in the limit as both $z \nearrow z_c$ and $x \rightarrow \infty$. The correlation length $\xi(z)$ is to be interpreted as the important length scale of the system. For $x \rightarrow \infty$ at fixed $z < z_c$, the two-point function is believed to obey the *Ornstein-Zernike* decay

$$G_z(0, x) \sim C_z |x|^{-(d-1)/2} e^{-|x|/\xi(z)}, \quad (2.1.1)$$

where C_z depends only on z , and strictly speaking the norm on the right side is equivalent to but not equal to the Euclidean norm. A proof of (2.1.1) for x on a coordinate axis will be given in Theorem 4.4.7. However the Ornstein-Zernike decay describes the behaviour of the two-point function on a length scale $|x| \gg \xi(z)$, and is not believed to be accurate on the important length scale where x is of the order of $\xi(z)$. Instead, the decay of the *critical* two-point function is considered to be fundamental on the scale of the correlation length, and we define a function $h(z; x)$ by

$$G_z(0, x) = \frac{1}{|x|^{d-2+\eta}} h(z; x). \quad (2.1.2)$$

The assumption now is that the important contribution to $h(z; x)$ will come from x of the order of the correlation length $\xi(z)$, and that we can write

$$G_z(0, x) \approx \frac{1}{|x|^{d-2+\eta}} g(|x|/\xi(z)), \quad (2.1.3)$$

for some universal function g of a single variable. The function g will be assumed to decay at infinity sufficiently rapidly that its product with any power of x is integrable, for example an exponential function.

Given (2.1.3), the following argument can be put forth in support of Fisher's relation $\gamma = (2 - \eta)\nu$. We assume that the main contribution to the susceptibility

$$\chi(z) = \sum_x G_z(0, x)$$

is due to x of the order of the correlation length, and that we may therefore substitute (2.1.3) into the sum over x . By definition of $\bar{\gamma}$ we then have

$$\begin{aligned} (z_c - z)^{-\bar{\gamma}} &\approx \sum_x \frac{g(|x|/\xi)}{|x|^{d-2+\eta}} \approx \int_0^\infty r^{1-\eta} g(r/\xi) dr \\ &= \text{const.} \xi^{2-\eta} \sim \text{const.} (z_c - z)^{-(2-\eta)\bar{\nu}}. \end{aligned} \quad (2.1.4)$$

This gives

$$\bar{\gamma} = (2 - \eta)\bar{\nu}. \quad (2.1.5)$$

It has already been argued in Section 1.3 that $\bar{\gamma} = \gamma$, and we will shortly argue that $\bar{\nu} = \nu$, which then gives $\gamma = (2 - \eta)\nu$.

The following is an alternate derivation of Fisher's relation which does not rely on (2.1.3). In the sum

$$G_{z_c}(0, x) = \sum_{N=0}^{\infty} c_N(0, x) \mu^{-N} \quad (2.1.6)$$

2.1. SCALING THEORY

we assume that $c_N(0, x)$ is significant only when $|x|$ is of the order of N^ν . There are about $N^{d\nu}$ such sites x , and we assume that an N -step self-avoiding walk is equally likely to end at any one of them, or in other words,

$$c_N(0, x) \approx c_N N^{-d\nu} \approx \mu^N N^{\gamma-1-d\nu}. \quad (2.1.7)$$

Restricting the sum in (2.1.6) to N between $c_1|x|^{1/\nu}$ and $c_2|x|^{1/\nu}$, for some positive constants c_1 and c_2 , and then using (2.1.7), gives

$$G_{z_c}(0, x) \approx \sum_{N=c_1|x|^{1/\nu}}^{c_2|x|^{1/\nu}} N^{\gamma-1-d\nu} \approx |x|^{(\gamma-d\nu)/\nu}. \quad (2.1.8)$$

This implies $-(d - 2 + \eta) = (\gamma - d\nu)/\nu$, which can be rewritten as $\gamma = (2 - \eta)\nu$.

Continuing in the spirit of the calculation leading to (2.1.5), we now argue that $\bar{\nu} = \nu_p = \nu$. For any $p \in (0, \infty)$,

$$\begin{aligned} \sum_x |x|^p G_z(0, x) &\approx \sum_x |x|^p \frac{g(|x|/\xi)}{|x|^{d-2+\eta}} \approx \int_0^\infty r^{p+1-\eta} g(r/\xi) dr \\ &= \text{const.} \xi^{p+2-\eta} \sim \text{const.} (z_c - z)^{-(p+2-\eta)\bar{\nu}}. \end{aligned} \quad (2.1.9)$$

Using the definition of ξ_p and (2.1.5), this gives

$$\xi_p(z)^p \approx (z_c - z)^{\bar{\gamma} - (2-\eta)\bar{\nu} - p\bar{\nu}} = (z_c - z)^{-p\bar{\nu}} \quad (2.1.10)$$

and hence $\bar{\nu} = \nu_p$. To show that $\bar{\nu} = \nu$, we first observe that by (1.3.20),

$$\sum_x |x|^2 G_z(0, x) \approx (z_c - z)^{-(2\nu+\gamma)}. \quad (2.1.11)$$

Comparing with (2.1.9), with $p = 2$, gives

$$(4 - \eta)\bar{\nu} = 2\nu + \gamma. \quad (2.1.12)$$

Now by (2.1.5) and the equality of $\bar{\gamma}$ and γ we conclude that $\bar{\nu} = \nu$.

We now turn to the hyperscaling relations (1.4.14) and (1.4.24). These cannot be derived from the scaling hypothesis (2.1.3) and require additional assumptions. Less numerical testing has been done of the hyperscaling relations than on the calculation of γ and ν , but both the Monte Carlo and series extrapolation computations which have been done are consistent with them.

The hyperscaling relation involving α_{sing} may at first glance seem somewhat surprising. It would perhaps seem natural to assume that the probability that an N -step self-avoiding walk ends at x would be proportional to

the characteristic volume $N^{-d\nu}$ in the limit as $N \rightarrow \infty$, as is the case for simple random walk. However this leads to the conclusion $\alpha_{sing} - \gamma - 1 = -d\nu$ rather than to the hyperscaling relation $\alpha_{sing} - 2 = -d\nu$. This incorrect argument fails to take into account the fact that for fixed x it is difficult for a long self-avoiding walk to return near to its starting point at the origin, and as the length of the walk goes to infinity x must be regarded as being close to the origin. The argument should be reasonable for x of the order of the typical length scale N^ν , and indeed this is what we assumed in the second derivation of Fisher's relation given above.

To obtain the hyperscaling relation $\alpha_{sing} - 2 = -d\nu$, we proceed as follows. First we assume that $c_N(0, x)$ will have the same scaling behaviour for any fixed x as $N \rightarrow \infty$, and consider the case $x = e$, with e a nearest neighbour of the origin. This assumption is mild in comparison with the assumptions we will make next. By adding an extra step to a walk ending at e we obtain a closed self-avoiding loop. Let $n = (N + 1)/2$. Then by summing over the position of the walk after n steps, and using symmetry, we have

$$c_N(0, e) = (2d)^{-1} \sum_x \sum_{\substack{\omega^{(1)} : \omega^{(1)}(n) = x \\ \omega^{(2)} : \omega^{(2)}(n) = x}} I[\omega^{(1)} \cap \omega^{(2)} = \{0, x\}]. \quad (2.1.13)$$

Here both of the self-avoiding walks $\omega^{(i)}$ begin at the origin and consist of n steps. We now make three assumptions. First, we assume that the main contribution to the above sum will be from x of the order of n^ν . Thus there are of the order of $n^{\nu d}$ relevant terms in the sum. Second, we assume that the effect of the avoidance constraint between $\omega^{(1)}$ and $\omega^{(2)}$ can be incorporated by replacing the quantity being summed over x by $c_N(0, x)^2$ multiplied by the square of the probability that two n -step self-avoiding walks beginning at the same point avoid each other; this probability is $c_{2n}/c_n^2 \sim 2^{\gamma-1} A^{-1} n^{1-\gamma}$. Here we use the square of this probability to account for the avoidance both near 0 and near x . Third, we assume that for x of the order of n^ν the probability that an n -step walk ends at x is of the order of the inverse of the characteristic volume $n^{\nu d}$, so that $c_n(0, x)$ is of the order of $\mu^n n^{\gamma-1} n^{-\nu d}$. With these three assumptions we have from (2.1.13) that

$$c_N(0, e) \approx n^{\nu d} [\mu^n n^{\gamma-1} n^{-\nu d}]^2 [n^{1-\gamma}]^2 = \mu^{2n} n^{-\nu d}. \quad (2.1.14)$$

Comparison with the definition of α_{sing} gives $\alpha_{sing} - 2 = -d\nu$.

We next turn to the hyperscaling relation (1.4.24) for Δ_4 , which is believed to hold only for $d \leq 4$. For simplicity, we take $N_1 = N_2 = n$ in the definition (1.4.21) of Δ_4 . To begin, we assume that since a self-avoiding

walk of length n goes a distance of about n^ν in each direction, such a self-avoiding walk will primarily lie in a hypercube of side n^ν and volume $n^{\nu d}$.

Given $0 \leq D \leq d$, consider a subset of the hypercubic lattice such that the cube of side R contains of the order of R^D points as $R \rightarrow \infty$. Such a subset can in some sense be thought of as being D -dimensional. From this point of view a long self-avoiding walk, which consists of n points in volume approximately equal to $n^{\nu d}$, is a $1/\nu$ -dimensional set. Typically two D -dimensional subsets will intersect in d dimensions if $2D \geq d$, but otherwise will not. This suggests that two n -step self-avoiding walks with a common point will typically have additional intersections if $d\nu \leq 2$, and typically will not if $d\nu > 2$.

Consider first the case of $d\nu > 2$, for which we have already seen in (1.4.26) that the hyperscaling relation fails. According to the values of ν given in Table 1.2, this inequality says $d > 4$. Two n -step self-avoiding walks lying in the cube of volume $n^{\nu d}$ will typically not intersect each other, and so there should be no overcounting in writing

$$c_{n,n} \approx c_n^2 n^2 \sim A^2 \mu^{2n} n^{2\gamma}. \quad (2.1.15)$$

Here the factor n^2 comes from choosing a point on each walk at which the two walks can be joined. By definition of Δ_4 , this gives $2\Delta_4 + \gamma - 2 = 2\gamma$, or $\Delta_4 = 1 + \gamma/2$. Using the mean-field value $\gamma = 1$ known to be correct for $d \geq 5$, we obtain $\Delta_4 = 3/2$ (which of course is consistent with the rigorous result for $d \geq 6$ obtained in Section 1.5).

We next consider the case $d\nu \leq 2$, which corresponds to $d \leq 4$. Here the factor of n^2 in (2.1.15) would overcount. Given one n -step walk, which will lie roughly within a cube of volume $n^{\nu d}$, a second n -step walk will typically intersect the first if it is started at any one of the $n^{\nu d}$ points in the cube. This leads to

$$c_{n,n} \approx c_n^2 n^{\nu d} \sim A^2 \mu^{2n} n^{2\gamma-2+\nu d}. \quad (2.1.16)$$

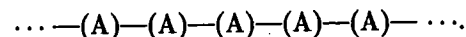
By definition of Δ_4 , this gives $2\Delta_4 + \gamma - 2 = 2\gamma - 2 + \nu d$, which simplifies to the hyperscaling relation $d\nu - 2\Delta_4 + \gamma = 0$.

2.2 Polymers

One of the most important applications of the self-avoiding walk is as a model for linear polymer molecules in chemical physics. In this section we shall briefly describe some aspects of this role, including a nonrigorous derivation of the "Flory values" for the critical exponent ν .

A *polymer* is a molecule that consists of many "monomers" (groups of atoms) joined together by chemical bonds. The *functionality* of a monomer

is the number of available chemical bonds that it has, i.e. the number of other monomers with which it must bond. If each monomer has functionality two, then a *linear* polymer is formed. If we denote a monomer by (A), then a linear polymer may be represented schematically as



One simple example is polyethylene, where each monomer is CH_2 (one carbon atom and two hydrogen atoms). The pattern terminates either by bonding with a monomer of functionality one, such as CH_3 , at each end, or else by closing on itself to form a “ring polymer”. When we speak of linear polymers, we shall be referring to the former. By way of contrast, if a polymer includes monomers of functionality three or more, then a *branched polymer* is formed; these are often modelled by lattice trees or lattice animals (see Section 5.5.1).

The preceding paragraph deals only with the topological structure of a polymer. Properties of its spatial configuration are no less important. Polymers can be very large; some linear polymers consist of more than 10^5 monomers. Thus the length scale of the entire polymer is macroscopic with respect to the length scale of the individual monomers. Consider a linear polymer consisting of $N+1$ monomers, and label the monomers $0, 1, \dots, N$ from one end to the other. Let $x(i) \in \mathbb{R}^3$ denote the location of the i -th monomer. Then the i -th (monomer-monomer) bond may be represented by the line segment joining $x(i-1)$ to $x(i)$. Typically, the length of each bond is essentially constant throughout the chain, as is the angle between each pair of consecutive monomer-monomer bonds. However, there is some rotational freedom for the i -th bond around the axis determined by the $(i-1)$ -th bond. In some cases, a reasonably good approximation may be obtained by allowing the rotational angle of the i -th bond around the $(i-1)$ -th bond to take on three different values, say 0° and $\pm 120^\circ$, perhaps with different probabilities (an angle of 0° means that the i -th, $(i-1)$ -th, and $(i-2)$ -th bonds all lie in one plane). These angles correspond to local configurations of minimal energy, and depend on the details of the monomers.

We see that one possible model for the spatial configuration of a linear polymer is simply a random walk in $\mathbb{R}^3$, and in fact this model is known as the *ideal polymer chain*. Alternatively, one can work with a lattice approximation, say a random walk on $\mathbb{Z}^3$. The model can be embellished by turning it into a Markov chain (or random walk with some finite memory), and it works reasonably well in some situations. However, there is a fundamental limitation of the ideal polymer chain, namely the *excluded volume* effect.

Two monomers cannot occupy the same position in space: the presence of a monomer at position x prohibits any other part of the polymer from getting too close to x , that is, other monomers are excluded from a certain volume of space. This is the excluded volume effect. When we take this effect into account, it becomes apparent that a self-avoiding walk is a more appropriate model for a linear polymer than is a random walk. The self-avoiding walk model is best for the case of a dilute polymer solution (where polymers are far apart, so that there is little interaction between distinct molecules) and a good solvent (which minimizes attractive forces between monomers).

We remark that there are some situations in which polymers really do behave ideally on large length scales, even though excluded volume effects are present. One is in a dense system (or “melt”) of many polymers, where monomers fill three-dimensional space uniformly and a given polymer interacts with many other monomers besides its own. Another is at certain values of temperature and solvent quality where roughly speaking the attractive forces between monomers exactly balance the excluded volume repulsion (the “ Θ point”). For more details, see the general polymer references listed in the Notes at the end of the chapter.

For the remainder of this section, we shall only discuss linear polymers in dilute solutions with good solvents. These are believed to be in the same “universality class” as the self-avoiding walk, which means in particular that they have the same critical exponents. For example, consider the *radius of gyration* of a polymer, which is the average distance of the monomers from the centre of mass of the polymer. The radius of gyration of polymers can be determined experimentally, for example from light scattering properties. For a polymer consisting of N monomers, the radius of gyration is expected to be asymptotic to DN^ν as $N \rightarrow \infty$, where D and ν are constants. The exponent ν is believed to be *universal*: it should be the same for all linear polymers (in dilute solution with good solvents), and for the self-avoiding walk as well. Moreover the exponent ν for the radius of gyration is believed to be the same as the critical exponent ν defined in (1.1.5) for the mean-square displacement, since polymers are expected to have only one macroscopic length scale. In contrast, the amplitude D is non-universal: it depends on microscopic details of the monomers and the solvent molecules.

The chemist Paul J. Flory developed an effective (but nonrigorous) method for computing the exponent ν [Flory (1949)]. We give a brief description of this method in general dimension d ; for simplicity, we ignore all multiplicative constants. (A more probabilistic description of the method will be given afterward.) Fix N and consider a linear polymer with $N+1$ monomers, represented by an N -step walk $\omega = (\omega(0), \dots, \omega(N))$ in $\mathbb{Z}^d$ (not

necessarily self-avoiding). Let L be the radius of gyration of ω , or any other "effective radius" of the walk. Then ω consists of $N + 1$ monomers (sites) spread through a box of volume L^d . Assuming uniformity, this gives a density of

$$\rho = \frac{N}{L^d} \quad (2.2.1)$$

monomers per unit volume. The repulsive energy per unit volume depends on the number of pairs of monomers per unit volume, which we approximate by ρ^2 . This is a "mean-field" approximation: it uses the assumption of uniformity very heavily, ignoring the strong correlations in the locations of consecutive monomers along the polymer. If we accept this approximation, then the total repulsive energy of the polymer is given by

$$E_{rep} = L^d \rho^2 = \frac{N^2}{L^d}. \quad (2.2.2)$$

Naturally the repulsive energy is lower for highly extended chains, i.e. large values of L .

Now consider the free energy F of the polymer of radius L , in the absence of the repulsion. This is given (up to constants) by (-1) times the entropy¹, and the entropy in turn is just the logarithm of the number of walks of radius L . Without repulsions, this can be found from the Gaussian behaviour of the ideal chain, as follows. Taking L now to denote the end-to-end distance and fixing $\omega(0) = 0$, we have

$$\Pr\{\omega(N) = x\} \approx N^{-d/2} \exp(-|x|^2/N) \quad (2.2.3)$$

for every $x \in \mathbb{Z}^d$, and hence

$$\Pr\{|\omega(N)| = L\} \approx \frac{L^{d-1}}{N^{d/2}} \exp(-L^2/N). \quad (2.2.4)$$

The total number of N -step walks is $(2d)^N$ in the nearest-neighbour case, so the free energy is

$$\begin{aligned} F &= -\log[(2d)^N \Pr\{|\omega(N)| = L\}] \\ &= -(d-1)\log L + \frac{L^2}{N} + \text{terms independent of } L. \end{aligned} \quad (2.2.5)$$

The term F may also be viewed as an "elastic energy" term, which prevents L from getting too large. The total energy of the polymer is now given by

¹In thermodynamics we have $F = U - TS$, where U is internal energy, T is temperature, and S is entropy. Here U depends on the number of monomers but not on L , so for our purposes it is constant and hence we ignore it.

the sum of the two energy terms (2.2.2) and (2.2.5):

$$E_{rep} + F = \frac{N^2}{L^d} + \frac{L^2}{N} - (d-1)\log L + K, \quad (2.2.6)$$

where K is independent of L . Now put $L = N^\nu$. Then the total energy (2.2.6) becomes

$$E_{rep} + F = N^{2-d\nu} + N^{2\nu-1} - \nu(d-1)\log N + K. \quad (2.2.7)$$

The value of ν that minimizes the energy (2.2.7) may be found by first equating the first two powers of N : solving $2 - d\nu = 2\nu - 1$ gives

$$\nu = \frac{3}{d+2}. \quad (2.2.8)$$

Substituting this back into (2.2.7), the first two terms become $N^{(4-d)/(d+2)}$, and these are the dominant terms if and only if $d < 4$. Therefore this argument predicts that (2.2.8) gives the correct value of ν whenever $d < 4$. When $d = 4$, this argument also predicts $\nu = 3/(4+2) = 1/2$ since this is the only value for which the first two terms of (2.2.7) remain bounded. However, when $d > 4$, any value of ν in the interval $[2/d, 1/2]$ keeps the first two terms of (2.2.7) bounded. Pushing this argument further suggests that we should take the largest value in this interval so as to minimize the $-\nu(d-1)\log N$ term in (2.2.7), obtaining $\nu = 1/2$ for $d > 4$. This answer makes sense: since ν equals $1/2$ in the ideal case, the addition of a repulsive energy term should not decrease ν below $1/2$, and so we conclude that $\nu = 1/2$ whenever $d > 4$.

To summarize, the above argument makes the following predictions for ν :

$$\nu_{Flory} = \begin{cases} 1 & \text{if } d = 1 \\ 3/4 & \text{if } d = 2 \\ 3/5 & \text{if } d = 3 \\ 1/2 & \text{if } d \geq 4. \end{cases} \quad (2.2.9)$$

These predictions are known as the *Flory values* for ν . As described in Section 1.1, they are known to be correct for $d = 1$ and $d \geq 5$, and they are believed to be correct for $d = 2$ and $d = 4$ as well. The Flory value for $d = 3$ is generally believed to be slightly too large: numerical and field theory calculations indicate that the actual value is probably close to 0.59 (some references are given in the Notes for Section 1.1). The success of Flory's argument is all the more remarkable when one realizes that it benefits greatly from the cancellation of two errors: both E_{rep} and F are greatly overestimated (see p. 46 of de Gennes (1979) for a brief discussion).

To conclude this section, we shall recast the Flory argument in a more probabilistic language. In (2.2.4) we calculated the probability that an N -step random walk ω (starting at the origin) has $|\omega(N)| = L$. Now let us estimate the probability that ω is self-avoiding given that $|\omega(N)| = L$. We shall write $L = N^\nu$ and choose ν to maximize this probability. As above, we assume that the $N+1$ sites of ω are spread uniformly through a box of volume L^d . Given that $\omega(0), \dots, \omega(k-1)$ are all distinct, the probability that $\omega(k)$ does not coincide with any one of the previous k sites is approximately $1 - kL^{-d}$ (this is the "mean-field" approximation). Hence the probability that ω is self-avoiding given that $|\omega(N)| = L$ is approximately

$$\prod_{k=1}^N (1 - kL^{-d}) \approx \exp \left(- \sum_{k=1}^N kL^{-d} \right) \approx \exp(-N^2/L^d). \quad (2.2.10)$$

Multiplying (2.2.10) by (2.2.4) yields

$$\Pr\{\omega \text{ is self-avoiding and } |\omega(N)| = L\} \approx \frac{L^{d-1}}{N^{d/2}} \exp \left[-\frac{L^2}{N} - \frac{N^2}{L^d} \right]. \quad (2.2.11)$$

To find the most likely value of L , we maximize the above probability for fixed N . Since the logarithm of this probability is just the negative of the total energy (2.2.6), we are again led to the Flory exponents.

2.3 The $N \rightarrow 0$ limit

In this section we describe a connection, discovered by de Gennes, between the self-avoiding walk and the spin systems of classical statistical mechanics: the self-avoiding walk can be considered to be a "zero-component" ferromagnet. Although this connection has not yet provided methods for obtaining rigorous results for the self-avoiding walk, it has been an important tool for physicists and has been used for example to compute the values for the critical exponents γ and ν for $d = 2, 3, 4$ given in (1.1.11) – (1.1.14). To make the discussion more self-contained, we first describe very briefly the basic set-up of spin models. The prototype of these models is the Ising model, and we begin with this fundamental model of ferromagnetism.

For simplicity we restrict attention to the hypercubic lattice $\mathbf{Z}^d$, although this is not essential. Let Λ denote the sites in $\mathbf{Z}^d$ which are in the cube $[-L, L]^d$, for $L \geq 1$. Eventually we will want to take the limit as $L \rightarrow \infty$. In the Ising model, a spin variable $S^{(x)}$ taking the value plus one or minus one is associated to each site $x \in \Lambda$. These spin variables interact

via a Hamiltonian

$$\mathcal{H} = - \sum_{(x,y)} S^{(x)} S^{(y)}, \quad (2.3.1)$$

where the sum represents the sum over all nearest-neighbour pairs of sites in Λ . The Hamiltonian represents the energy of a spin configuration (choice of ± 1 for each spin), and is lowest when neighbouring spins agree. The expected value of any function F of the spins in Λ is then given by

$$\langle F \rangle = \frac{1}{Z} E(F e^{-\beta \mathcal{H}}), \quad (2.3.2)$$

where the expectation E on the right side is with respect to the product of the Bernoulli measures assigning probability one-half to each of the possible values ± 1 for the spin variables, and the *partition function*

$$Z = E(e^{-\beta \mathcal{H}}) \quad (2.3.3)$$

is a normalization factor. The nonnegative parameter β corresponds to inverse temperature. The partition function and expectations depend on the volume Λ , but to simplify the notation no subscripts Λ will be used to keep track of this.

An important example is $F = S^{(0)} S^{(x)}$, the product of the values of the spins at the origin and at x . For any finite volume Λ it follows from the symmetry of the Hamiltonian under the global spin flip, in which each spin is multiplied by minus one, that $\langle S^{(y)} \rangle = 0$ for any site $y \in \Lambda$. Hence the *two-point function* $\langle S^{(0)} S^{(x)} \rangle$ represents the correlation between the spins at the origin and at x . It follows from the fact that the two-point function lies in the compact interval $[0, 1]$ that there is a subsequence of volumes tending to infinity such that the limit of the two-point function exists along the subsequence. (The same subsequence can be used for all x by a diagonal argument.) In fact it can be shown using correlation inequalities that the infinite volume limit of the two-point function exists, without recourse to subsequences. The infinite volume limit is often referred to as the *thermodynamic limit*. Here we are using *free boundary conditions*, in which spins on the inside boundary of Λ interact only with their nearest neighbours inside Λ . It is known that in the thermodynamic limit, for high temperatures (or in other words for low β) the two-point function decays exponentially as $|x| \rightarrow \infty$. The inverse of the decay rate defines a correlation length $\xi(\beta)$. For dimensions $d \geq 2$ there is a critical value β_c (corresponding to the Curie point) such that the correlation length diverges to infinity as $\beta \nearrow \beta_c$. This corresponds to the onset of long range order.

Associated with the critical point β_c , a number of critical exponents can be defined which are analogous to the exponents defined for the self-avoiding walk. For example a critical exponent, known as ν as for the

self-avoiding walk, defines the power law according to which the correlation length diverges:

$$\xi(\beta) \sim \text{const.}(\beta_c - \beta)^{-\nu} \text{ as } \beta \nearrow \beta_c. \quad (2.3.4)$$

The susceptibility is defined by

$$\chi(\beta) = d^{-1} \sum_x \langle S^{(0)} \cdot S^{(x)} \rangle, \quad (2.3.5)$$

for the infinite volume theory with $\beta < \beta_c$. The susceptibility diverges as $\beta \nearrow \beta_c$ and the power law at which the divergence takes place defines a critical exponent γ :

$$\chi(\beta) \sim \text{const.}(\beta_c - \beta)^{-\gamma} \text{ as } \beta \nearrow \beta_c. \quad (2.3.6)$$

These qualitative analogies between the critical behaviours of spin systems and the self-avoiding walk can be made more quantitative. For this we need to introduce a generalization of the Ising model, known as the N -vector or $O(N)$ model. In this generalization the Ising spins are replaced by spins taking values in the N -dimensional sphere of radius $\sqrt{N}$, for some positive integer N , and the Hamiltonian becomes

$$\mathcal{H} = - \sum_{\langle x, y \rangle} S^{(x)} \cdot S^{(y)}, \quad (2.3.7)$$

where the dot product is the usual Euclidean one. The two-point function for the N -vector model is then defined in finite volume as for the Ising model, with the change that now the single spin distribution is the uniform measure on $S(N, \sqrt{N})$, where

$$S(n, r) = \{(a_1, \dots, a_n) \in \mathbb{R}^n : a_1^2 + \dots + a_n^2 = r^2\} \quad (2.3.8)$$

is the sphere of radius r in $\mathbb{R}^n$. For $N = 1$ this is just the Ising model. For $N \geq 2$ the N -vector model also has a critical point, and shares many common features with the Ising model (although the change from discrete to continuous symmetry group introduces new elements). Critical exponents can be defined, which will in general depend on N as well as on the dimension d . In a manner to be described in more detail below, the N -vector model can be defined in the limit as $N \rightarrow 0$, and this limit gives the self-avoiding walk. The N -vector model can be analyzed, at least non-rigorously, using renormalization methods, and this analysis yields values for the critical exponents in which N appears as a parameter which may be assigned values other than positive integers. Taking $N = 0$ in the expression for the critical exponents then gives values which are believed to

correspond to the self-avoiding walk exponents, and indeed the values given for two dimensions in (1.1.11) and (1.1.12) were obtained by Nienhuis in this way. The self-avoiding walk exponents for three dimensions and the logarithmic corrections for $d = 4$ can be arrived at similarly.

To take the $N \rightarrow 0$ limit, consider a fixed finite volume Λ with free boundary conditions (i.e. we consider only sites in Λ in the sum defining the Hamiltonian, and do not take the infinite volume limit). Our aim is to show that the two-point function for the N -vector model converges to that of the self-avoiding walk, i.e. for any $\beta \geq 0$ and for any fixed i, j and sites x, y ,

$$\lim_{N \rightarrow 0} \langle S_i^{(x)} S_j^{(y)} \rangle = \delta_{i,j} \sum_{\omega: x \rightarrow y} \beta^{|\omega|} = \delta_{i,j} G_\beta(x, y), \quad (2.3.9)$$

where the subscript i denotes the i -th spin component and the sum is over all self-avoiding walks (in Λ) of any length, from x to y . In the process, it will be necessary to define what is meant by the limit on the left side of (2.3.9), since the N -vector model two-point function has only been defined when N is a positive integer. To begin with some notation, for a function F of the spins in Λ and for N a positive integer, we write

$$\langle F \rangle = \frac{1}{Z} E(F e^{-\beta \mathcal{H}}) = \frac{1}{Z} \int F e^{-\beta \mathcal{H}} d\mu_N, \quad (2.3.10)$$

where $d\mu_N$ denotes the product over the spins of uniform measures on $S(N, \sqrt{N})$, and the partition function Z is the normalization

$$Z = E(e^{-\beta \mathcal{H}}) = \int e^{-\beta \mathcal{H}} d\mu_N. \quad (2.3.11)$$

To obtain (2.3.9) it will be argued that

$$\lim_{N \rightarrow 0} Z = \lim_{N \rightarrow 0} E(e^{-\beta \mathcal{H}}) = 1 \quad (2.3.12)$$

and

$$\lim_{N \rightarrow 0} E(S_i^{(x)} S_j^{(y)} e^{-\beta \mathcal{H}}) = \delta_{i,j} \sum_{\omega: x \rightarrow y} \beta^{|\omega|}. \quad (2.3.13)$$

The analysis will not proceed by extending the definitions of the expectations whose limits are being taken in the above two equations to positive real values of N , and then taking the limit in the strict mathematical sense. Rather, we will show that a certain plausible interpretation of the limit leads to (2.3.12) and (2.3.13); thus our arguments do not lead to these equations as rigorous mathematical statements.

Both (2.3.12) and (2.3.13) will be obtained in the same way. The first step is to expand the exponential in a power series:

$$e^{-\beta\mathcal{H}} = \prod_{\langle x,y \rangle} \exp[\beta S^{(x)} \cdot S^{(y)}] = \prod_{\langle x,y \rangle} \sum_{m_{xy}=0}^{\infty} \frac{\beta^{m_{xy}}}{m_{xy}!} (S^{(x)} \cdot S^{(y)})^{m_{xy}}. \quad (2.3.14)$$

Then we label the nearest-neighbour (undirected) bonds of Λ by $b_1, \dots, b_B$, and for each bond b_α label one of its endpoints b_α^- and the other b_α^+ . In this notation (2.3.14) can be written

$$e^{-\beta\mathcal{H}} = \sum_{m_1, \dots, m_B=0}^{\infty} \frac{\beta^{\sum \alpha m_\alpha}}{\prod_\alpha m_\alpha!} \prod_\alpha (S^{(b_\alpha^-)} \cdot S^{(b_\alpha^+)})^{m_\alpha}. \quad (2.3.15)$$

Hence Z or the two-point function be computed in terms of expectations of products of powers of spin components. Such expectations can be evaluated using the following lemma.

Lemma 2.3.1 Fix an integer $N \geq 1$. Let $S = (S_1, \dots, S_N)$ denote a vector which is uniformly distributed on $S(N, \sqrt{N})$. Given nonnegative integers $k_1, \dots, k_N$,

$$E(S_1^{k_1} \dots S_N^{k_N}) = \begin{cases} \frac{2\Gamma(\frac{N+2}{2}) \prod_{i=1}^N \Gamma(\frac{k_i+1}{2})}{\pi^{N/2} \Gamma(\frac{k_1+\dots+k_N+N}{2})} N^{(k_1+\dots+k_N-2)/2} & \text{all } k_i \text{ even} \\ 0 & \text{otherwise,} \end{cases}$$

where Γ denotes the Gamma function.

Proof. The lemma is clearly true for $N = 1$, so we fix $N \geq 2$. Suppose $U = (U_1, \dots, U_N)$ is uniformly distributed on the sphere $S(N, 1)$. Using the fact that $\Gamma(\frac{N+2}{2}) = \frac{N}{2} \Gamma(\frac{N}{2})$, it suffices to show for all $k = 1, \dots, N$ that if the integers m_i are all even then

$$E(U_1^{m_1} U_2^{m_2} \dots U_k^{m_k}) = \frac{\Gamma(\frac{N}{2}) \prod_{i=1}^k \Gamma(\frac{m_i+1}{2})}{\pi^{k/2} \Gamma(\frac{m_1+\dots+m_k+N}{2})}, \quad (2.3.16)$$

and that this expectation is equal to 0 if any m_i is an odd integer. The latter follows by symmetry. We will prove the former by induction on k .

For $k = 1$, we use the fact [proved in Watson (1983), p.44] that the marginal density of U_i is

$$\frac{\Gamma(\frac{N}{2})}{\pi^{1/2} \Gamma(\frac{N-1}{2})} (1-a^2)^{(N-3)/2}, \quad -1 \leq a \leq 1. \quad (2.3.17)$$

It then follows from the identity

$$\int_0^1 t^c (1-t)^d dt = \frac{\Gamma(c+1)\Gamma(d+1)}{\Gamma(c+d+2)} \quad (2.3.18)$$

(for $c, d > -1$) that

$$E(U_1^m) = 0 \quad \text{if } m \text{ is odd, and} \quad (2.3.19)$$

$$E(U_1^m) = \frac{\Gamma(\frac{N}{2}) \Gamma(\frac{m+1}{2})}{\pi^{1/2} \Gamma(\frac{N+m}{2})} \quad \text{if } m \text{ is even} \quad (2.3.20)$$

which gives (2.3.16) for $k = 1$.

Suppose now that $k > 1$ and assume that (2.3.16) is true for $k-1$. Conditioned on $U_1 = a$, the distribution of $(U_2, \dots, U_N)$ is uniform on the set $S(N-1, (1-a^2)^{1/2})$. The inductive hypothesis then gives the conditional expectation

$$\begin{aligned} E(U_2^{m_2} \dots U_k^{m_k} | U_1 = a) \\ = (1-a^2)^{(m_2+\dots+m_k)/2} \frac{\Gamma(\frac{N-1}{2}) \prod_{i=2}^k \Gamma(\frac{m_i+1}{2})}{\pi^{(k-1)/2} \Gamma(\frac{m_2+\dots+m_k+N-1}{2})}. \end{aligned} \quad (2.3.21)$$

Inserting this into

$$\begin{aligned} E(U_1^{m_1} U_2^{m_2} \dots U_k^{m_k}) \\ = \int_{-1}^{+1} a^{m_1} E(U_2^{m_2} \dots U_k^{m_k} | U_1 = a) \frac{\Gamma(\frac{N}{2})}{\pi^{1/2} \Gamma(\frac{N-1}{2})} (1-a^2)^{(N-3)/2} da \end{aligned}$$

then gives the desired result (2.3.16). $\square$

We now use Lemma 2.3.1 to define what we mean by the limit as $N \rightarrow 0$ of expectations like those in the statement of the lemma. It follows from the lemma that for any positive integer N and any index i ,

$$E(S_i^2) = 1;$$

in fact this can be seen more easily by symmetry and the fact that $E(S_1^2 + \dots + S_N^2) = N$. We will therefore assert, by way of definition, that

$$\lim_{N \rightarrow 0} E(S_i^2) = 1. \quad (2.3.22)$$

Also, we will define the limit as $N \rightarrow 0$ of any expectation as in the statement of the lemma to be zero if $k_1 + \dots + k_N > 2$. This is consistent with the result of the lemma; e.g. if two k_i 's equal 2 and the others are 0, then

$$E(S_i^2 S_j^2) = \frac{2\Gamma(\frac{N+2}{2}) (\frac{1}{2})^2}{\Gamma(\frac{N+4}{2})} N, \quad (2.3.23)$$

which converges to zero as $N \rightarrow 0$. According to this definition,

$$\lim_{N \rightarrow 0} E(S_1^{k_1} \dots S_N^{k_N}) = \begin{cases} 1 & \text{all } k_l = 0, \text{ or one } k_l = 2 \text{ and } k_j = 0, j \neq l \\ 0 & \text{otherwise.} \end{cases} \quad (2.3.24)$$

Consider now the partition function

$$Z = \sum_{m_1, \dots, m_B=0}^{\infty} \frac{\beta^{\sum_{\alpha} m_{\alpha}}}{\prod_{\alpha} m_{\alpha}!} E \left(\prod_{\alpha} (S^{(b_{\alpha}^-)} \cdot S^{(b_{\alpha}^+)})^{m_{\alpha}} \right). \quad (2.3.25)$$

Equation (2.3.24) provides a means of extending Z to $N = 0$, by taking the limit as $N \rightarrow 0$ termwise in the above sum. A graphical interpretation of the sum in (2.3.25) can be obtained by associating to each term in the sum a graph whose edges are given by m_i undirected edges joining the endpoints of the bond b_i . It then follows from (2.3.24) that any term whose corresponding graph has a vertex from which other than two or zero edges emanate will approach zero in the limit as $N \rightarrow 0$. This can be seen by considering a specific example. Consider the graph consisting of the four nearest-neighbour edges $\{z, x\}, \{z, y\}, \{z, w\}$. The expectation arising from the corresponding term in (2.3.25) is

$$\sum_{i,j,k,l} E \left(S_i^{(y)} S_i^{(z)} S_j^{(w)} S_j^{(z)} S_k^{(x)} S_k^{(z)} S_l^{(x)} S_l^{(z)} \right), \quad (2.3.26)$$

where the sum is over spin components. Since spins at different sites are independent, the expectation in the above sum factors into a product of four expectations. The factor corresponding to the site z is

$$E \left(S_i^{(z)} S_j^{(z)} S_k^{(z)} S_l^{(z)} \right), \quad (2.3.27)$$

which will go to zero in the limit as $N \rightarrow 0$, for any choice of i, j, k, l , by (2.3.24). There is further N -dependence arising from the number of terms in the sum over spin components i, j, k, l , but this will be interpreted as only helping to drive the limit to zero.

The relevant graphs in the limit are therefore the graphs consisting of a finite number (possibly zero) of nonintersecting self-avoiding polygons in Λ , where the degenerate polygons consisting of two edges linking a pair of nearest-neighbour sites are allowed possibilities. The graph with no edges corresponds to the term in the sum with all $m_i = 0$, and contributes an amount 1. A two-edge polygon with nearest-neighbour vertices x, y contributes an amount

$$\frac{\beta^2}{2} E \left([S^{(x)} \cdot S^{(y)}]^2 \right). \quad (2.3.28)$$

Since an expectation involving an odd power of $S_i^{(x)}$ is zero, this is equal to

$$\frac{\beta^2}{2} \sum_{i=1}^N E \left([S_i^{(x)} S_i^{(y)}]^2 \right) = \frac{\beta^2}{2} N \quad (2.3.29)$$

and hence does not contribute in the limit.

A nondegenerate polygon, in other words a polygon consisting of at least four bonds, contributes an amount

$$\beta^k E \left((S^{(y_1)} \cdot S^{(y_2)})(S^{(y_2)} \cdot S^{(y_3)}) \dots (S^{(y_r)} \cdot S^{(y_1)}) \right) \quad (2.3.30)$$

where y_l is a neighbour of y_{l+1} for each $l = 1, \dots, r-1$, y_r is a neighbour of y_1 , and $y_1, \dots, y_r$ are distinct. This expression is equal to

$$\beta^k \sum_{i=1}^N E \left((S_i^{(y_1)} S_i^{(y_2)} \dots S_i^{(y_r)})^2 \right) = \beta^k N, \quad (2.3.31)$$

which also converges to 0 as $N \rightarrow 0$. We are thus led to conclude that

$$\lim_{N \rightarrow 0} Z = 1. \quad (2.3.32)$$

For the two-point function the analysis is similar. We would like to compute the limit as $N \rightarrow 0$ of the expectation

$$\sum_{m_1, \dots, m_B=0}^{\infty} \frac{\beta^{\sum_{\alpha} m_{\alpha}}}{\prod_{\alpha} m_{\alpha}!} E \left(S_i^{(x)} S_j^{(y)} \prod_{\alpha} (S^{(b_{\alpha}^-)} \cdot S^{(b_{\alpha}^+)})^{m_{\alpha}} \right). \quad (2.3.33)$$

For $x = y$ the limit of the above expression is equal to 1 by an analysis similar to that used to analyze the partition function. Suppose now that $x \neq y$. Again there is a correspondence between the terms in the sum and graphs on Λ , but now there can be a nonzero contribution to the limit only from those graphs in which exactly one edge emanates from each of the vertices x and y , and either two or zero edges emanate from every other vertex. Such a graph must consist of a self-avoiding walk from x to y together with a finite number of (possibly degenerate) self-avoiding polygons. The contribution to the limit from the polygons is equal to zero as it is for the partition function. The contribution due to the self-avoiding walk with vertices $(x, v_1, \dots, v_{k-1}, y)$ is

$$\beta^k E \left(S_i^{(x)} (S^{(x)} \cdot S^{(v_1)})(S^{(v_1)} \cdot S^{(v_2)}) \dots (S^{(v_{k-1})} \cdot S^{(y)}) S_j^{(y)} \right). \quad (2.3.34)$$

This expression is equal to $\beta^k \delta_{i,j}$, since upon expanding the dot products everything has expectation 0 except $(S_i^{(x)} S_i^{(y_1)} \dots S_i^{(y_{k-1})} S_i^{(y)})^2$. Since there is one such term for every self-avoiding walk from x to y ,

$$\lim_{N \rightarrow 0} \langle S_i^{(x)} S_j^{(y)} \rangle = \delta_{i,j} \sum_{\omega: x \rightarrow y} \beta^{|\omega|} = \delta_{i,j} G_\beta(x, y). \quad (2.3.35)$$

This correspondence of two-point functions is responsible for the general belief that the critical exponents γ and ν also correspond in the $N \rightarrow 0$ limit.

We end this section with a nonrigorous discussion of the equality in the $N \rightarrow 0$ limit of the self-avoiding walk critical exponent α_{sing} , defined in (1.4.13), and the critical exponent for the singular part of the specific heat of the N -vector model. To distinguish between these two exponents we shall denote the latter by α_s .

To define the specific heat we first introduce the expected energy per unit volume, which is given by

$$\mathcal{E}(\beta) = \langle \sum_{(0,x)} S^{(0)} \cdot S^{(x)} \rangle = 2dN \langle S_i^{(0)} S_i^{(e)} \rangle, \quad (2.3.36)$$

for any fixed i and any nearest neighbour e of the origin. The prefactor $2dN$ is irrelevant as far as the behaviour of $\mathcal{E}(\beta)$ near the critical point is concerned, so we introduce

$$u(\beta) = \langle S_i^{(0)} S_i^{(e)} \rangle. \quad (2.3.37)$$

The specific heat is defined as the rate of change of the energy with respect to temperature β^{-1} , i.e.

$$C(\beta) = -\beta^2 \frac{du}{d\beta}. \quad (2.3.38)$$

Typically the specific heat either diverges as β increases to β_c , or there is a nonnegative integer M such that it has M but not $M+1$ derivatives at β_c^- , with

$$C(\beta) \approx \sum_{j=0}^M \frac{C^{(j)}(\beta_c)}{j!} (\beta - \beta_c)^j + (\beta_c - \beta)^{-\alpha_s}, \quad (2.3.39)$$

for some exponent $-\alpha_s \in (M, M+1]$. In principle both α_s and M can depend on N . Here we are using the symbol $\approx$ which indicates a crude correspondence between the right and left sides; in particular in the correction term we are dropping sign and constant factors, and also possible logarithmic factors which can be expected to be present when $-\alpha_s$ is an integer.

We also allow $M = -1$ in (2.3.39), with the empty sum interpreted as zero, to deal simultaneously with both finite and infinite $C(\beta_c)$. For $M = -1$, requiring that the energy $u(\beta)$ remains bounded as $\beta \nearrow \beta_c$ implies that $-\alpha_s \in (-1, 0]$.

In view of (2.3.38), the behaviour (2.3.39) of the specific heat suggests that

$$u(\beta) \approx \sum_{j=0}^{M+1} \frac{u^{(j)}(\beta_c)}{j!} (\beta - \beta_c)^j + (\beta_c - \beta)^{-\alpha_s+1}. \quad (2.3.40)$$

Assuming that the form of the above relation persists in the limit as $N \rightarrow 0$, from (2.3.37) and (2.3.35) we obtain

$$G_\beta(0, e) = \sum_{n=1}^{\infty} c_n(0, e) z^n \approx \sum_{j=0}^{M+1} g_j (\beta - \beta_c)^j + (\beta_c - \beta)^{-\alpha_s+1}, \quad (2.3.41)$$

where

$$g_j = \frac{1}{j!} \frac{d^j}{d\beta^j} G_\beta(0, e) \Big|_{\beta_c} = \sum_{n=1}^{\infty} c_n(0, e) \binom{n}{j} \beta_c^{n-j}. \quad (2.3.42)$$

The exponent α_s in (2.3.41) is interpreted as the $N \rightarrow 0$ limit of an N -dependent exponent. Our goal now is to argue that this limiting value of α_s is equal to α_{sing} . We further assume that in (2.3.41) M is the largest integer such that g_{M+1} is finite, so that $g_{M+2} = \infty$. Assuming now that $c_n(0, e) \approx n^{\alpha_{sing}-2} \mu^n = n^{\alpha_{sing}-2} \beta_c^{-n}$, g_j will be finite if and only if $\alpha_{sing} - 2 + j < -1$. Thus we have $-\alpha_{sing} \in (M, M+1]$, which is consistent with the restriction $-\alpha_s \in (M, M+1]$ in (2.3.39).

Writing $\beta = \beta_c e^{-t}$, so that $\beta_c - \beta \sim \beta_c t$, we have

$$G_\beta(0, e) - \sum_{j=0}^{M+1} g_j (\beta - \beta_c)^j \approx \sum_{n=1}^{\infty} n^{\alpha_{sing}-2} \left[e^{-nt} - \sum_{j=0}^{M+1} \binom{n}{j} (-t)^j \right]. \quad (2.3.43)$$

Approximating the sum over n on the right side by an integral and then making the change of variables $y = xt$, the right side of (2.3.43) is given approximately by

$$\begin{aligned} & \int_1^{\infty} x^{\alpha_{sing}-2} \left[e^{-xt} - \sum_{j=0}^{M+1} \binom{x}{j} (-t)^j \right] dx \\ &= t^{-\alpha_{sing}+1} \int_t^{\infty} y^{\alpha_{sing}-2} \left[e^{-y} - \sum_{j=0}^{M+1} \binom{y/t}{j} (-t)^j \right] dy. \end{aligned}$$

As $t \rightarrow 0$, the right side behaves like

$$t^{-\alpha_{sing}+1} \int_t^\infty y^{\alpha_{sing}-2} \left[e^{-y} - \sum_{j=0}^{M+1} \frac{1}{j!} (-y)^j \right] dy. \quad (2.3.44)$$

In view of the fact that $-\alpha_{sing} \in (M, M+1]$, the above integral is convergent both for large y and for $y \rightarrow 0$ (apart from a logarithmic divergence as $y \rightarrow 0$ when $-\alpha_{sing} = M+1$). The integral is clearly nonzero, since the quantity in square brackets in the integrand is of the same sign for all positive y , by Taylor's Theorem with remainder. Hence the overall behaviour is $t^{-\alpha_{sing}+1}$. Comparing now with (2.3.41), we conclude that $\alpha_s = \alpha_{sing}$.

2.4 Notes

Section 2.1. Scaling theory is discussed in many theoretical physics texts on critical phenomena, for example Amit (1984), and we shall make no attempt here to refer to the original literature.

Section 2.2. Some general references on polymers which elaborate on the topics mentioned here include Flory (1971), de Gennes (1979), Doi and Edwards (1986), and des Cloiseaux and Jannink (1990). A readable survey is given in Flory's 1974 Nobel lecture [Flory (1976)]. Whittington (1982) discusses several additional topics in the statistical mechanics of polymers and self-avoiding walks.

Flory (1949) originally discussed only the three-dimensional case of the argument presented in this section. The extension to other dimensions was first observed by Fisher (1969). There are many other arguments which derive the Flory exponents; for example, see Edwards (1965), Freed (1981), and Bouchaud and Georges (1989).

Section 2.3. For a general introduction to rigorous results for spin systems, see for example Fernández, Fröhlich and Sokal (1992), Thompson (1988), Glimm and Jaffe (1987), Ellis (1985), Ruelle (1969). More physics-oriented accounts are given in for example Itzykson and Drouffe (1989), Parisi (1988), Amit (1984).

The fact that the $N \rightarrow 0$ limit of the N -vector model gives the self-avoiding walk was first observed in de Gennes (1972); see also de Gennes (1979). The use of Lemma 2.3.1 in deriving the $N \rightarrow 0$ limit appears to be new. Other approaches can be found in Aragão de Carvalho, Caracciolo and Fröhlich (1983), Halley and Dasgupta (1983), Domb (1976), and Bowers and McKerrell (1973). The calculation of critical exponents for the

two-dimensional N -vector model, and corresponding identification of the exponents for $N = 0$, was carried out in Nienhuis (1982); see also Nienhuis (1984) and Nienhuis (1987). For $d = 3$, the $N = 0$ critical exponents are calculated in Le Guillou and Zinn-Justin (1989). Logarithmic corrections in four dimensions were computed in Larkin and Khmel'Nitskii (1969), Wegner and Riedel (1973) and Brezin, Le Guillou and Zinn-Justin (1973).

There is an intimate relation between spin systems and interacting random walks of various types (going far beyond the $N \rightarrow 0$ limit). This was emphasized by Symanzik (1969), and developed further in Brydges, Fröhlich and Spencer (1982). A detailed account of the random walk representations of spin systems is given in Fernández, Fröhlich and Sokal (1992).

Chapter 3

Some combinatorial bounds

3.1 The Hammersley-Welsh method

As was mentioned in Section 1.1, there is still no rigorous proof of the finiteness of the critical exponent γ for the number of self-avoiding walks [see Equation (1.1.4)] in dimensions two, three and four. The best rigorous upper bounds on c_N/μ^N are essentially of the form $\exp(O(N^p))$ for some constant $0 < p < 1$. It is a major open problem to replace this bound by a polynomial in N . We remark that subadditivity (Section 1.2) by itself gives no information about such subexponential behaviour.

Theorem 3.1.1 below, with its elegant proof, is due to Hammersley and Welsh (1962). Although this result (which holds for all $d \geq 2$) has subsequently been improved for $d > 2$, after three decades it remains the best rigorous upper bound on c_N in two dimensions. Improved bounds for $d > 2$ can be obtained, with considerably more work, using an extension of the Hammersley-Welsh method. These improved bounds, which are given in Theorem 3.3.1, remain the best available in three and four dimensions. In five or more dimensions, entirely different methods have been used to prove that c_N/μ^N is asymptotically constant (and hence bounded); these methods will be described in Chapter 6.

Theorem 3.1.1 *Let $d \geq 2$. For any constant $B > \pi(2/3)^{1/2}$, there exists an $N_0(B)$ independent of d such that*

$$c_N \leq \mu^{N+1} e^{BN^{1/2}} \text{ for all } N \geq N_0. \quad (3.1.1)$$

The proof relies on bridges (see Definition 1.2.4), and yields as a bonus the fact that $\mu_{\text{Bridge}} = \mu$. In particular, it uses the fact that walks are “subadditive” [Equation (1.2.3)] while bridges are “superadditive” [Equation (1.2.15)] and plays these two off against each other. The basic idea is that every self-avoiding walk can be “unfolded” into a bridge, and that this transformation is at most $\exp(O(N^{1/2}))$ -to-one. Before we give the details, we require a few definitions, as well as a classical theorem of number theory which we will quote without proof.

Definition 3.1.2 An N -step half-space walk is an N -step self-avoiding walk ω whose first components satisfy the inequality

$$\omega_1(0) < \omega_1(i) \text{ for all } i = 1, \dots, N.$$

The number of N -step half-space walks starting at the origin is denoted h_N . By convention, $h_0 = 1$.

In particular, every bridge is a half-space walk.

Definition 3.1.3 The span of an N -step self-avoiding walk ω is

$$\max_{0 \leq j \leq N} \omega_1(j) - \min_{0 \leq j \leq N} \omega_1(j).$$

The number of N -step half-space walks (respectively, bridges) starting at the origin and having span A is denoted $h_{N,A}$ (respectively, $b_{N,A}$).

Note that $h_{N,0}$ is 1 if $N = 0$ and is 0 otherwise.

Theorem 3.1.4 For each integer $A \geq 1$, let $P_D(A)$ denote the number of partitions of A into distinct integers (i.e. the number of ways to write $A = A_1 + \dots + A_k$ where $A_1 > \dots > A_k$). Then

$$\log P_D(A) \sim \pi \left(\frac{A}{3} \right)^{1/2} \text{ as } A \rightarrow \infty. \quad (3.1.2)$$

This theorem is proved in Hardy and Ramanujan (1917).

The following proposition contains the first part of the proof of Theorem 3.1.1, in which half-space walks are “unfolded” into bridges by a sequence of reflections.

Proposition 3.1.5 For every $N \geq 1$,

$$h_N \leq P_D(N) b_N, \quad (3.1.3)$$

where $P_D(N)$ is defined in Theorem 3.1.4.

Proof. Let $N \geq 1$, and let ω be an N -step half-space walk that starts at the origin. Let $n_0 = 0$. For each $j = 1, 2, \dots$, recursively define $A_j(\omega)$ and $n_j(\omega)$ so that

$$A_j = \max_{n_{j-1} < i \leq N} (-1)^j (\omega_1(n_{j-1}) - \omega_1(i))$$

and n_j is the largest value of i for which this maximum is attained. The recursion is stopped at the smallest integer k such that $n_k = N$; this means that $A_{k+1}(\omega)$ and $n_{k+1}(\omega)$ are not defined. (See Figure 3.1.) Observe that

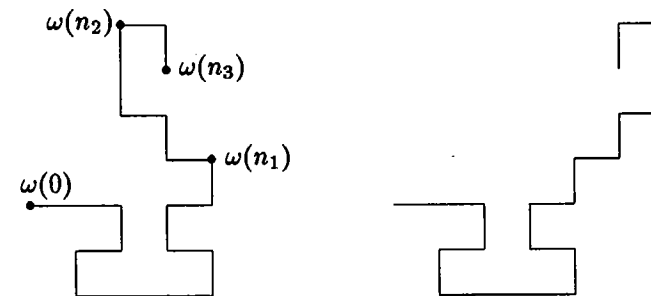


Figure 3.1: A half-space walk ω in $H_{20}[4, 2, 1]$ and the transformed walk ω' in $H_{20}[6, 1]$.

$A_1(\omega)$ is the span of ω ; in general, $A_{j+1}(\omega)$ is the span of the self-avoiding walk $(\omega(n_j), \dots, \omega(N))$, which is either a half-space walk or the reflection of one. Moreover, each of the subwalks $(\omega(n_j), \dots, \omega(n_{j+1}))$ is either a bridge or the reflection of one. Also observe that $A_1 > A_2 > \dots > A_k > 0$.

For every decreasing sequence of k positive integers $a_1 > a_2 > \dots > a_k > 0$, let $H_N[a_1, \dots, a_k]$ be the set of N -step half-space walks ω with $\omega(0) = 0$ and $A_1(\omega) = a_1, \dots, A_k(\omega) = a_k$, and $n_k(\omega) = N$ (and hence $A_{k+1}(\omega)$ is not defined). Note that in particular $H_N[a]$ is the set of N -step bridges of span a .

Given an N -step half-space walk ω , define a new N -step walk ω' as follows: for $0 \leq i \leq n_1(\omega)$, define $\omega'(i) = \omega(i)$; and for $n_1(\omega) < i \leq N$, define $\omega'(i)$ to be the reflection of the point $\omega(i)$ in the hyperplane $x_1 = A_1(\omega)$. Observe that if ω is in $H_N[a_1, a_2, \dots, a_k]$, then ω' is in $H_N[a_1 + a_2, a_3, \dots, a_k]$; moreover, this transformation is one-to-one, so

$$|H_N[a_1, a_2, \dots, a_k]| \leq |H_N[a_1 + a_2, a_3, \dots, a_k]|.$$

Therefore, summing over all finite integer sequences $a_1 > \dots > a_k > 0$,

$$h_N = \sum |H_N[a_1, \dots, a_k]|$$

$$\begin{aligned} &\leq \sum |H_N[a_1 + \cdots + a_k]| \\ &= \sum b_{N, a_1 + \cdots + a_k}, \end{aligned}$$

which tells us that

$$h_N \leq \sum_{A=1}^N P_D(A) b_{N,A}. \quad (3.1.4)$$

Since $P_D(A) \leq P_D(N)$ for $A \leq N$, it follows from (3.1.4) that

$$h_N \leq P_D(N) \sum_{A=1}^N b_{N,A} = P_D(N) b_N, \quad (3.1.5)$$

which proves the proposition. $\square$

We now complete the proof of the Hammersley-Welsh bound on c_N . The idea is to split each self-avoiding walk into two half-space walks, and then to use Proposition 3.1.5.

Proof of Theorem 3.1.1. Fix $B > \pi(2/3)^{1/2}$, and choose $\epsilon > 0$ so that $B - \epsilon > \pi(2/3)^{1/2}$. By Theorem 3.1.4, there exists a constant K such that

$$P_D(A) \leq K \exp \left[(B - \epsilon)(A/2)^{1/2} \right] \text{ for all } A. \quad (3.1.6)$$

Given an arbitrary n -step self-avoiding walk ω , let $M = \min_i \omega_1(i)$ and let m be the largest i such that $\omega_1(i) = M$. Then $(\omega(m), \dots, \omega(n))$ is a half-space walk, as is

$$(\omega(m) - (1, 0, 0, \dots, 0), \omega(m), \omega(m-1), \dots, \omega(0)).$$

Using this decomposition, as well as Proposition 3.1.5, the inequality $b_i b_j \leq b_{i+j}$ [from (1.2.15)], (3.1.6), and the inequality $x^{1/2} + y^{1/2} \leq (2x + 2y)^{1/2}$, we obtain

$$\begin{aligned} c_n &\leq \sum_{m=0}^n h_{n-m} h_{m+1} \\ &\leq \sum_{m=0}^n b_{m+1} b_{n-m} P_D(m+1) P_D(n-m) \\ &\leq b_{n+1} \sum_{m=0}^n K^2 \exp \left((B - \epsilon) \left[\left(\frac{m+1}{2} \right)^{1/2} + \left(\frac{n-m}{2} \right)^{1/2} \right] \right) \\ &\leq b_{n+1} (n+1) K^2 \exp \left[(B - \epsilon)(n+1)^{1/2} \right] \end{aligned} \quad (3.1.7)$$

for all n . Therefore, there exists an $N_0(B)$ (independent of d) such that

$$c_n \leq b_{n+1} e^{Bn^{1/2}} \text{ for all } n \geq N_0. \quad (3.1.8)$$

Since $b_{n+1} \leq \mu^{n+1}$ [Equation (1.2.17)], Theorem 3.1.1 is now proven. $\square$

Corollary 3.1.6 Let B be as in Theorem 3.1.1. Then, for all sufficiently large N ,

$$\mu^{N-1} e^{-BN^{1/2}} \leq b_N \leq \mu^N. \quad (3.1.9)$$

In particular,

$$\lim_{N \rightarrow \infty} (b_N)^{1/N} = \mu. \quad (3.1.10)$$

Proof. The right inequality of (3.1.9) is just (1.2.17); the left inequality comes from the bound $\mu^n \leq c_n$ [recall (1.2.10)] and from (3.1.8) (with n replaced by $N - 1$). Equation (3.1.10) is then immediate. $\square$

Definition 3.1.7 The generating function for the number of bridges is denoted B_z and is given by

$$B_z = \sum_{N=0}^{\infty} b_N z^N.$$

Equation (3.1.10) says that the radius of convergence of B_z is $z_c = \mu^{-1}$. The following corollary says that B_z actually diverges at $z = z_c$.

Corollary 3.1.8

$$\lim_{z \nearrow z_c} B_z = +\infty;$$

that is,

$$B_{z_c} = \sum_{N=1}^{\infty} b_N \mu^{-N} = +\infty. \quad (3.1.11)$$

Proof. The proof of Theorem 3.1.1 shows that every N -step half-space walk may be decomposed into a finite sequence of bridges $\{\omega^{(i)}\}$ having spans A_i and lengths m_i , where $A_1 > A_2 > \dots$ and $\sum m_i = N$; moreover, the sequence of bridges uniquely determines the original half-space walk. Therefore for $N \geq 1$

$$h_N \leq \sum \left(\prod_{i=1}^k b_{m_i, A_i} \right) \quad (3.1.12)$$

where the sum is over all integers $k \geq 1$, all integers $A_1 > \dots > A_k > 0$, and all integers $m_1, \dots, m_k \geq 1$ that sum to N . Consequently, for $z > 0$

$$\sum_{N=0}^{\infty} h_N z^N \leq \prod_{A=1}^{\infty} \left(1 + \sum_{m=1}^{\infty} b_{m,A} z^m \right);$$

this can be seen by comparing z^N terms on both sides and using (3.1.12). Combining this inequality with the elementary inequality $1 + x \leq e^x$, we find

$$\sum_{N=0}^{\infty} h_N z^N \leq \exp \left(\sum_{A=1}^{\infty} \sum_{n=1}^{\infty} b_{n,A} z^n \right) = \exp(B_z - 1).$$

This and the first inequality of (3.1.7) imply

$$\sum_{N=0}^{\infty} c_N z^N \leq z^{-1} \left(\sum_{n=0}^{\infty} h_n z^n \right)^2 \leq z^{-1} e^{2(B_z - 1)}. \quad (3.1.13)$$

By Equation (1.3.6), the leftmost term of (3.1.13) diverges at $z = z_c$, hence so does the rightmost term. This proves the corollary. $\square$

We remark that the above proof gives an explicit bound on the rate of divergence of B_z : indeed, combining (3.1.13) with (1.3.6) yields

$$B_z \geq 1 + \frac{1}{2} \log \frac{z_c z}{z_c - z} \quad \text{for } 0 < z < z_c. \quad (3.1.14)$$

3.2 Self-avoiding polygons

Intuitively, a *self-avoiding polygon* may be thought of as a simple (i.e. non-self-intersecting) closed curve embedded in the lattice, with neither starting point nor orientation specified. The precise definition is as follows.

Definition 3.2.1 Let N be an integer greater than 2. An N -step self-avoiding polygon is a set $\mathcal{P}$ of N nearest-neighbour bonds with the following property: there exists a corresponding $(N-1)$ -step self-avoiding walk ω having $|\omega(N-1) - \omega(0)| = 1$ such that $\mathcal{P}$ consists of precisely the bond joining $\omega(N-1)$ to $\omega(0)$ and the $N-1$ bonds joining $\omega(i-1)$ to $\omega(i)$ ($i = 1, \dots, N-1$).

Observe that ω is not uniquely determined by $\mathcal{P}$; in fact, each N -step self-avoiding polygon has precisely $2N$ corresponding self-avoiding walks (there are N choices of starting point and two choices of orientation). However,

3.2. SELF-AVOIDING POLYGONS

no $(N-1)$ -step self-avoiding walk corresponds to more than one N -step polygon.

We want to count self-avoiding polygons by ignoring translations and only counting different shapes; thus, in $\mathbb{Z}^2$, there should be only one 4-step self-avoiding polygon (a unit square) and two 6-step self-avoiding polygons (rectangles, one being a 90° rotation of the other). This leads us to the following definition.

Definition 3.2.2 Two N -step self-avoiding polygons are said to be equivalent up to translation if there is a vector v in $\mathbb{R}^d$ such that translation by v defines a one-to-one correspondence from the set of bonds of one polygon to the set of bonds of the other polygon. Also, we denote by q_N the number of distinct equivalence classes up to translation of N -step self-avoiding polygons.

Thus, if e is one of the $2d$ nearest neighbours of the origin in $\mathbb{Z}^d$, then the observations following Definition 3.2.1 tell us that

$$2Nq_N = 2dc_{N-1}(0, e) \quad (3.2.1)$$

for every $N > 2$ (recall from Section 1.3 that $c_n(x, y)$ is the number of n -step self-avoiding walks from x to y). In particular, for $d = 2$, we have $q_4 = 1$, $c_3(0, e) = 2$, $q_6 = 2$, and $c_5(0, e) = 6$. Observe that $q_N = 0$ for every odd N .

Two self-avoiding polygons can be concatenated to form a larger self-avoiding polygon. The procedure is clear in two dimensions (see Figure 3.2): join a "rightmost" bond of one to a "leftmost" bond of the other. In higher

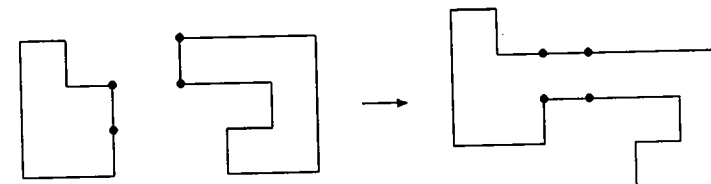


Figure 3.2: Concatenation of a 10-step polygon and a 14-step polygon to produce a 24-step polygon in $\mathbb{Z}^2$. The dots are the endpoints of the bonds that are changed during the concatenation.

dimensions, however, the procedure is slightly more involved, because such a pair of edges need not be parallel. In general, the concatenation effectively occurs in a $(d-1)$ -dimensional hyperplane, and so there will be an additional

factor of $d - 1$ to account for the number of possible orientations of the "leftmost" bond. The correct form of the subadditivity relation for polygons is the following.

Theorem 3.2.3 For even integers $M, N \geq 4$,

$$\frac{q_M q_N}{d-1} \leq q_{N+M} \quad (3.2.2)$$

and

$$q_N \leq q_{N+2}. \quad (3.2.3)$$

Proof. We prove (3.2.2) first; the proof of (3.2.3), which is similar, will follow. First, we define the *lexicographic ordering* on $\mathbb{Z}^d$, as follows. We say that $(a_1, \dots, a_d) < (b_1, \dots, b_d)$ if for some j (with $1 \leq j \leq d$) we have: $a_i = b_i$ whenever $1 \leq i < j$, and $a_j < b_j$. For even integers $N \geq 4$, let $Q[N]$ be the set of N -step self-avoiding polygons whose lexicographically smallest point is the origin. Then $Q[N]$ has exactly q_N members.

For each $i = 1, \dots, d$, let $e^{(i)}$ be the neighbour of the origin with $e_i^{(i)} = 1$ and $e_j^{(i)} = 0$ for $j \neq i$. For $i = 2, \dots, d$ and for even $M \geq 4$, let $Q_i[M]$ be the set of M -step self-avoiding polygons that lie in the half-space $x_1 \geq 0$ and that contain the bond joining the origin to $e^{(i)}$. Then $Q[M]$ is contained in the union of $Q_2[M], \dots, Q_d[M]$, and so, by symmetry,

$$|Q_2[M]| = \dots = |Q_d[M]| \geq \frac{q_M}{d-1}. \quad (3.2.4)$$

Choose an arbitrary N -step polygon $\mathcal{P}$ in $Q[N]$, and let p be its lexicographically largest point. There are two values of i ($1 \leq i \leq d$) such that $\mathcal{P}$ contains the bond joining p to $p - e^{(i)}$; let I be the larger of these two values. (In particular, we have $I \geq 2$.) Then let $\mathcal{Q}$ be an arbitrary self-avoiding polygon in $Q_I[M]$.

We now concatenate $\mathcal{P}$ and $\mathcal{Q}$. First translate $\mathcal{Q}$ by the vector $p - e^{(I)} + e^{(1)}$ (so the resulting polygon lies in the half-space $x_1 \geq p_1 + 1$ and contains the bond joining $p - e^{(I)} + e^{(1)}$ to $p + e^{(1)}$). Then take all of the bonds in the translated $\mathcal{Q}$ *except* the bond joining $p - e^{(I)} + e^{(1)}$ to $p + e^{(1)}$, and all of the bonds of $\mathcal{P}$ *except* the bond joining p to $p - e^{(I)}$, and also take the two bonds that join $p - e^{(I)}$ to $p - e^{(I)} + e^{(1)}$ and p to $p + e^{(1)}$. Since $\mathcal{P}$ is contained in the half-space $x_1 \leq p_1$, the result is a self-avoiding polygon in $Q[N+M]$. Conversely, given an $(N+M)$ -step polygon constructed in this fashion, we can reconstruct $\mathcal{P}$ and $\mathcal{Q}$, because the N sites with smallest first coordinate are precisely the points of $\mathcal{P}$. (Of course, not every polygon in $Q[N+M]$ can be obtained by such a concatenation.)

Since there were q_N ways to choose $\mathcal{P}$, and at least $q_M/(d-1)$ ways to choose $\mathcal{Q}$ given $\mathcal{P}$ [by (3.2.4)], inequality (3.2.2) follows immediately.

Finally we prove (3.2.3). Choose $\mathcal{P}$, p , and I as above. Then remove the bond joining p to $p - e^{(I)}$ from $\mathcal{P}$, and add the three bonds of the walk $(p, p + e^{(1)}, p + e^{(1)} - e^{(I)}, p - e^{(I)})$ to $\mathcal{P}$. The result is a self-avoiding polygon in $Q[N+2]$ from which $\mathcal{Q}$ can be unambiguously determined as above. This proves (3.2.3). $\square$

Now let $a_1 = 0$ and $a_n = -\log(q_{2n}/(d-1))$ for $n \geq 2$. Then Theorem 3.2.3 says that $\{a_n\}_{n \geq 1}$ is a subadditive sequence. Therefore Lemma 1.2.2 implies that $\lim_{n \rightarrow \infty} (q_{2n}/(d-1))^{1/2n}$ exists and equals some number $\mu_{\text{Polygon}} \leq \mu$, and that

$$q_N \leq (d-1)(\mu_{\text{Polygon}})^N \quad (3.2.5)$$

for all even $N > 2$. In fact, $\mu_{\text{Polygon}} = \mu$; this will be a corollary of the next theorem, independent of Theorem 3.2.3.

Theorem 3.2.4 Let e be a nearest neighbour of the origin in $\mathbb{Z}^d$. There exists a constant K , depending only on the dimension d , such that for every integer $M \geq 1$,

$$c_{2M+1}(0, e) \geq KM^{-d-2}(b_M)^2. \quad (3.2.6)$$

Proof. For a point x in $\mathbb{Z}^d$, let $B[M, x]$ denote the set of M -step bridges which begin at the origin and end at x , and let $|B[M, x]|$ denote the number of bridges in this set.

Consider a point x for which $B[M, x]$ is not empty, and let ω and v be bridges in $B[M, x]$ (not necessarily different). See Figure 3.3. Choose any vector $\mathbf{v} \equiv \mathbf{v}(x)$ in $\mathbb{R}^d$ which is orthogonal to the line containing 0 and x . Let i (respectively j) be chosen from among those values of $\{0, 1, \dots, M\}$ that maximize (respectively, minimize) the dot product $\omega(i) \cdot \mathbf{v}$ (respectively, $v(j) \cdot \mathbf{v}$). Define

$$\begin{aligned} \bar{\omega} &= (\omega(i), \dots, \omega(M), \omega(1) + \omega(M), \dots, \omega(i) + \omega(M)), \\ \bar{v} &= (v(j), \dots, v(M), v(1) + v(M), \dots, v(j) + v(M)). \end{aligned}$$

It is not hard to check that $\bar{\omega}$ and $\bar{v}$ are both self-avoiding walks (since ω and v were bridges), that $\bar{\omega}(M) - \bar{\omega}(0)$ and $\bar{v}(M) - \bar{v}(0)$ both equal x , and that

$$\bar{\omega}(0) \cdot \mathbf{v} = \bar{\omega}(M) \cdot \mathbf{v} \geq \bar{\omega}(k) \cdot \mathbf{v}$$

$$\bar{v}(0) \cdot \mathbf{v} = \bar{v}(M) \cdot \mathbf{v} \leq \bar{v}(k) \cdot \mathbf{v}$$

for all $k = 0, \dots, M$. To interpret these inequalities, think of two hyperplanes orthogonal to $\mathbf{v}$, one passing through $\bar{\omega}(0)$ and the other through $\bar{v}(0)$; then $\bar{\omega}$ and $\bar{v}$ lie on opposite sides of their respective hyperplanes.

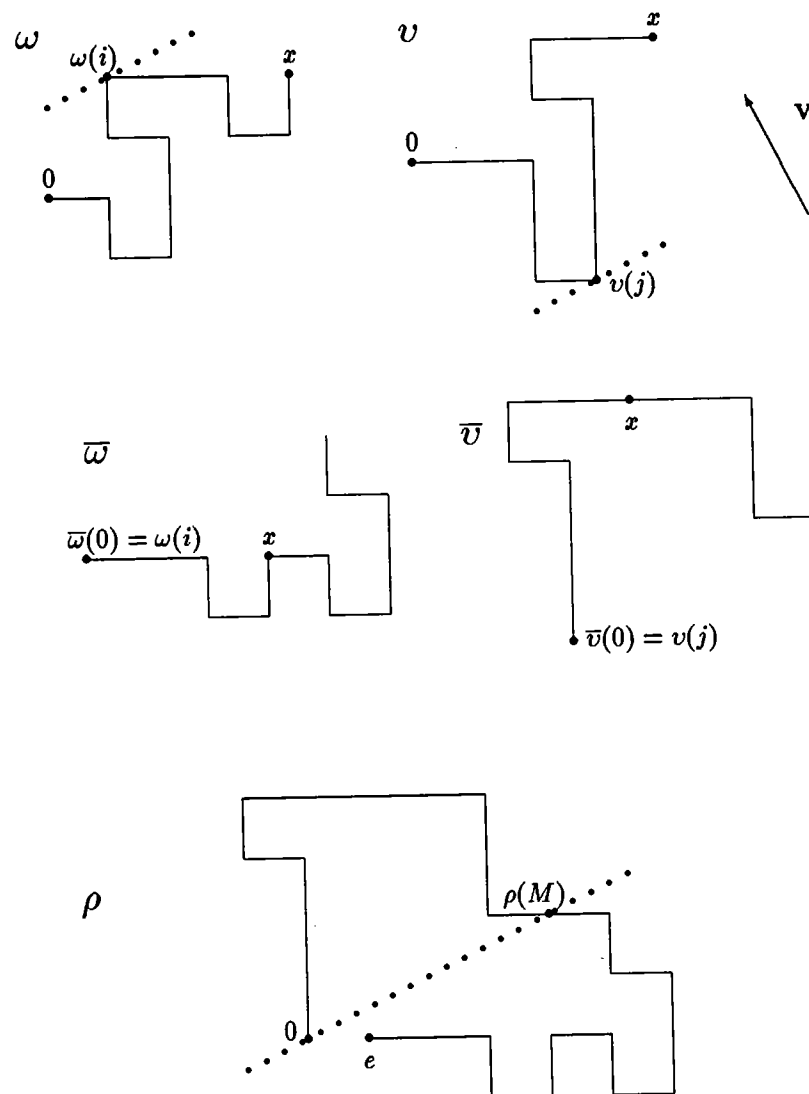


Figure 3.3: Proof of Theorem 3.2.4. Here $M = 12$. Top: the M -step bridges ω and v , and the vector $\mathbf{v}$. Middle: the derived walks $\bar{\omega}$ and $\bar{v}$. Bottom: the $(2M+1)$ -step walk ρ . The dotted lines are orthogonal to $\mathbf{v}$.

Now let e be a nearest neighbour of the origin such that $e \cdot \mathbf{v} < 0$. Let ϱ be the $(2M+1)$ -step walk starting at the origin and consisting of $\bar{v}$, followed by one step in the e direction, followed by the reversal of $\bar{\omega}$; that is,

$$\varrho(k) = \begin{cases} \bar{v}(k) - \bar{v}(0) & \text{for } 0 \leq k \leq M, \\ \bar{\omega}(2M+1-k) - \bar{\omega}(0) + e & \text{for } M+1 \leq k \leq 2M+1. \end{cases}$$

Then ϱ is a self-avoiding walk since the hyperplane with normal vector $\mathbf{v}$ that passes through the origin separates the first $M+1$ points of ϱ from the last $M+1$. Also, $\varrho(2M+1) - \varrho(0) = e$ and $\varrho(M) = x$.

Given a self-avoiding walk ϱ that has been constructed as above, we could reconstruct the original bridges ω and v if we only knew i and j . There are $M+1$ possible values for each of i and j . Therefore, if $\mathcal{S}$ denotes the set of $(2M+1)$ -step self-avoiding walks ϱ with $\varrho(0) = 0$ and $|\varrho(2M+1)| = 1$, then the number of walks in $\mathcal{S}$ having $\varrho(M) = x$ is at least $|B[M, x]|^2 / (M+1)^2$. Since there are fewer than $M(2M+1)^{d-1}$ values of x for which $|B[M, x]| > 0$, it follows from the above argument and the Schwarz inequality that

$$|\mathcal{S}| \geq \frac{\sum_x |B[M, x]|^2}{(M+1)^2} \geq \frac{(\sum_x |B[M, x]|)^2}{(M+1)^2 M (2M+1)^{d-1}}. \quad (3.2.7)$$

The theorem is a direct consequence of (3.2.7). $\square$

Corollary 3.2.5 *There exists a constant C depending only on the dimension d such that*

$$\mu^{2M} e^{-CM^{1/2}} \leq c_{2M+1}(0, e) \leq \frac{2(M+1)(d-1)}{d} \mu^{2M+2} \quad (3.2.8)$$

for all $M \geq 1$. In particular we have

$$\mu_{\text{Polygon}} = \lim_{n \rightarrow \infty} (q_{2n})^{1/2n} = \mu. \quad (3.2.9)$$

Proof. The first inequality of (3.2.8) is a direct consequence of Theorem 3.2.4 and Equation (3.1.9) (the constant C can absorb all factors of polynomial order). The second inequality follows from (3.2.1), (3.2.5), and the obvious bound $\mu_{\text{Polygon}} \leq \mu$. Finally, Equation (3.2.9) follows immediately from Equations (3.2.8) and (3.2.1). $\square$

Remark. It is possible to prove that $\mu_{\text{Polygon}} = \mu_{\text{Bridge}}$ directly, without using the results of Section 3.1. This can be done using Theorem 3.2.4 to prove $\mu_{\text{Polygon}} \geq \mu_{\text{Bridge}}$, and the bound $q_N \leq d(d-1)b_N$, which follows from Proposition 8.1.2, to prove $\mu_{\text{Polygon}} \leq \mu_{\text{Bridge}}$.

Corollary 3.2.6 *Let $\{x^{(N)}\}$ be a sequence of sites in $\mathbb{Z}^d \setminus \{0\}$ such that $|x^{(N)}| = o(N)$ as $N \rightarrow \infty$. To avoid trivialities, we also assume that $\|x^{(N)}\|_1$ has the same parity as N . Then*

$$\lim_{N \rightarrow \infty} \left(c_N(0, x^{(N)}) \right)^{1/N} = \mu. \quad (3.2.10)$$

In particular, for every $x \neq 0$, the two-point function $G_z(0, x)$ has the same radius of convergence $z_c = \mu^{-1}$ as the susceptibility $\chi(z)$.

Proof. For fixed N , let ϕ be a fixed self-avoiding walk from the origin to $x^{(N)}$ of length $\|x^{(N)}\|_1$ (or possibly $\|x^{(N)}\|_1 + 2$) whose lexicographically largest point, p , is neither 0 nor $x^{(N)}$. Let L_N be the length of ϕ . Choose $I \geq 2$ so that the bond joining p and $p - e^{(I)}$ is a bond of ϕ . We now consider concatenation of ϕ and self-avoiding polygons Q in $Q_I[N - L_N]$, as in the fourth paragraph of the proof of Theorem 3.2.3. In detail: Given such a polygon Q , translate it by the vector $p - e^{(I)} + e^{(1)}$. Then take all of the bonds in the translated Q except the bond joining $p - e^{(I)} + e^{(1)}$ to $p + e^{(1)}$, and all of the bonds of ϕ except the bond joining p to $p - e^{(I)}$, and also take the two bonds that join $p - e^{(I)}$ to $p - e^{(I)} + e^{(1)}$ and p to $p + e^{(1)}$. Since ϕ is contained in the half-space $x_1 \leq p_1$, and since the translated polygon lies in the half-space $x_1 \geq p_1 + 1$, the result determines an N -step self-avoiding walk from 0 to $x^{(N)}$. We conclude from (3.2.4) that

$$c_N(0, x^{(N)}) \geq \frac{q_{N-L_N}}{d-1}. \quad (3.2.11)$$

Since $L_N = o(N)$, the result now follows immediately from Corollary 3.2.5 and the trivial bound $c_N(0, x) \leq c_N$. $\square$

We remark that for fixed x (i.e. $x^{(N)}$ independent of N), the lower bound (3.2.11) can be improved by a factor of order N ; see Proposition 7.4.4.

3.3 Kesten's bound on c_N

In this section, we shall prove the following upper bound on the number of self-avoiding walks:

Theorem 3.3.1 *Let $d \geq 2$. Then there exists a constant Q , depending only on d , such that for every $N \geq 2$*

$$c_N \leq \mu^N \exp[QN^{2/(d+2)} \log N]. \quad (3.3.1)$$

3.3. KESTEN'S BOUND ON c_N

As we observed in Section 1.1, this is the best bound that is known rigorously in three and four dimensions. In two dimensions, it is not quite as good as the Hammersley-Welsh bound (Theorem 3.1.1), while above four dimensions we know that $c_N \sim \text{const.} \mu^N$ (see Section 6.1).

This theorem first appeared in Kesten (1964). However, that paper only presented a proof of the weaker bound

$$c_N \leq \mu^N \exp[QN^{2/(d+1)} \log N]. \quad (3.3.2)$$

The proof of Theorem 3.3.1 builds on the proof of (3.3.2) and draws on the ideas of the Hammersley-Welsh argument. The full proof was not given in Kesten (1964) because it was hoped that someone would find a better bound, but almost thirty years later this has still not come to pass for $d = 3, 4$. The proof of Theorem 3.3.1 that we present here is due to Kesten (private communication).

To begin with, observe that by the first inequality of (3.1.7) and the inequality $x^a + y^a \leq 2^{1-a}(x+y)^a$ for $0 < a < 1$ and $x, y \geq 0$, it suffices to prove the same bound for half space walks, i.e. that there exists a Q depending only on d such that

$$h_n \leq \mu^n \exp[Qn^{2/(d+2)} \log n]. \quad (3.3.3)$$

for every $n \geq 2$. Therefore we shall work with half-space walks for much of the proof. We first need an extension of Definitions 3.1.2 and 3.1.3.

Definition 3.3.2 *For integers N and S , let $h_{N,S}^* = \sum_{i=0}^S h_{N,i}$ be the number of half-space walks starting at the origin and having span at most S .*

Consider an integer $n \geq 1$. If ω is an n -step half-space walk that starts at the origin, then let $K(\omega)$ denote the span of ω and let $I(\omega)$ be the largest value of i such that $\omega_1(i) = K(\omega)$. Since $|\omega| > 0$, both $K(\omega)$ and $I(\omega)$ are nonzero. Observe that the first $I(\omega)$ steps of ω is a bridge of span $K(\omega)$, and the remainder of ω is (the reflection of) a half-space walk whose span is less than $K(\omega)$. Since there are at most n^2 possibilities for the pair $I(\omega)$ and $K(\omega)$, there exist integers $i[0]$ and $k[0]$ in $\{1, \dots, n\}$ such that the number of half-space walks ω having $I(\omega) = i[0]$ and $K(\omega) = k[0]$ is at least $n^{-2}h_n$. Therefore the above decomposition shows that

$$h_n \leq n^2 b_{i[0], k[0]} h_{n-i[0], k[0]}^*. \quad (3.3.4)$$

At this point, we shall state a lemma which is crucial for the proof of Theorem 3.3.1. Its proof will be deferred to the end of the section. To help the reader appreciate the role of the lemma, we shall show how the weaker bound (3.3.2) may be obtained as an immediate corollary of the lemma and (3.3.4).

Lemma 3.3.3 Let k, l , and m be strictly positive integers, and let B be a real number satisfying $0 < B < 1$. Let $V = (m^{1-B}l)^{1/d}$. Then there exists a constant D , depending only on the dimension d , such that

$$b_{l,k} h_{m,k}^* \leq \mu^{m+l+dV} [D(m+l)]^{12m^B+3dV}. \quad (3.3.5)$$

Corollary 3.3.4 Let $d \geq 2$. There exists a constant Q , depending only on d , such that (3.3.2) holds for every $n \geq 2$.

Proof. As explained prior to (3.3.3), it suffices to prove (3.3.2) with c_n replaced by h_n . Let $B = 2/(d+1)$, and let $n \geq 2$. If $i[0] = n$, then the result follows from (3.3.4) together with the basic relations $b_n \leq \mu^n$ and $h_{0,k}^* = 1$ for every k . Therefore, assume that $i[0] < n$. By (3.3.4) and Lemma 3.3.3, we have

$$h_n \leq n^2 \mu^{n+dV} [Dn]^{12n^B+3dV}, \quad (3.3.6)$$

where

$$V = ((n - i[0])^{1-B} i[0])^{1/d} \leq n^{(2-B)/d}. \quad (3.3.7)$$

Since $(2-B)/d = 2/(d+1) = B$, we see that

$$h_n \leq \mu^n (\mu^d)^{n^B} [Dn]^{(12+3d)n^B+2}, \quad (3.3.8)$$

and the result follows. $\square$

We now proceed with the proof of Theorem 3.3.1. The idea is to iterate (3.3.4) until certain auxiliary conditions are satisfied, and then to make some estimates and apply Lemma 3.3.3. Fix real numbers A and B in the interval $(0, 1)$, and fix an integer $n \geq 1$. (As we shall see by the end of the proof, we are specifically interested in the values $A = d/(d+2)$ and $B = 2/(d+2)$.)

We shall now give a procedure for defining an integer $u \geq 0$ and integers $i[0], \dots, i[u], k[0], \dots, k[u] > 0$ (all depending on A and n) having certain properties. We have already seen how to define $i[0]$ and $k[0]$. Next, if $n^2 b_{i[0],k[0]} > \mu^{i[0]}$ and $i[0] < n^A$, or if $i[0] = n$, then set $u = 0$ and stop; otherwise we reapply the decomposition of (3.3.4) with n replaced by $n - i[0]$ to choose $i[1]$ and $k[1]$ from $\{1, \dots, n - i[0]\}$ such that

$$h_{n-i[0],k[0]}^* \leq (n - i[0])^2 b_{i[1],k[1]} h_{n-i[0]-i[1],k[1]}^* \quad (3.3.9)$$

(notice that $h_{n-i[0],k[0]}^* \leq h_{n-i[0]}$). Then (3.3.4) and (3.3.9) imply that

$$h_n \leq n^2 b_{i[0],k[0]} (n - i[0])^2 b_{i[1],k[1]} h_{n-i[0]-i[1],k[1]}^*. \quad (3.3.10)$$

We now repeat this procedure inductively. Suppose that $i[j]$ and $k[j]$ have already been defined but $i[j+1]$ and $k[j+1]$ have not yet been defined. If

$$(n - i[0] - \dots - i[j-1])^2 b_{i[j],k[j]} > \mu^{i[j]} \quad (3.3.11)$$

and

$$i[j] < (n - i[0] - \dots - i[j-1])^A, \quad (3.3.12)$$

or if $i[0] + \dots + i[j] = n$, then set u equal to this value of j and stop the procedure. Otherwise, choose $i[j+1]$ and $k[j+1]$ from $\{1, \dots, n - i[0] - \dots - i[j]\}$ such that

$$\begin{aligned} h_{n-i[0]-\dots-i[j],k[j]}^* \\ \leq (n - i[0] - \dots - i[j])^2 b_{i[j+1],k[j+1]} h_{n-i[0]-\dots-i[j+1],k[j+1]}^*. \end{aligned} \quad (3.3.13)$$

Thus we end up with the inequality

$$\begin{aligned} h_n \leq n^2 b_{i[0],k[0]} (n - i[0])^2 b_{i[1],k[1]} \dots \\ \times (n - i[0] - \dots - i[u-1])^2 b_{i[u],k[u]} h_{n-i[0]-\dots-i[u],k[u]}^*. \end{aligned} \quad (3.3.14)$$

Let $\mathcal{I}$ be the set of j 's in $\{0, 1, \dots, u-1\}$ with the property that (3.3.11) holds (if $u = 0$, then $\mathcal{I}$ is the empty set). Therefore (3.3.12) fails for these values of j , i.e.

$$i[j] \geq (n - i[0] - \dots - i[j-1])^A \quad \text{for every } j \text{ in } \mathcal{I}. \quad (3.3.15)$$

If $0 \leq j < u$ and j is not in $\mathcal{I}$, then the reverse inequality of (3.3.11) holds (with $\leq$), while if j is in $\mathcal{I}$ or if j equals u , then we have the simple inequality

$$(n - i[0] - \dots - i[j-1])^2 b_{i[j],k[j]} \leq n^2 \mu^{i[j]} \quad (3.3.16)$$

(from $b_i \leq \mu^i$). Applying these inequalities to (3.3.14) yields

$$h_n \leq n^{2|\mathcal{I}|+2} \mu^{i[0]+\dots+i[u]} h_{n-i[0]-\dots-i[u],k[u]}^*. \quad (3.3.17)$$

Next we claim that there exists a constant C depending on A but not on n such that

$$|\mathcal{I}| \leq Cn^{1-A}. \quad (3.3.18)$$

To see this, for each integer $a \geq 0$ we let $\mathcal{I}_a$ denote the subset of integers j in $\mathcal{I}$ with the property that

$$n2^{-a} \geq n - i[0] - \dots - i[j-1] \geq n2^{-a-1}. \quad (3.3.19)$$

If we can show that

$$|\mathcal{I}_a| \leq 1 + (n2^{-a-1})^{1-A} \quad (3.3.20)$$

for every $a \geq 0$, then the claim (3.3.18) will follow from

$$|\mathcal{I}| \leq \sum_{a=0}^{\log_2 n} |\mathcal{I}_a| \leq 1 + \log_2 n + \sum_{a=0}^{\infty} (n2^{-a-1})^{1-A}. \quad (3.3.21)$$

If $|\mathcal{I}_a| \leq 1$, then (3.3.20) is trivial. Otherwise let f_a and F_a denote the smallest and largest members of $\mathcal{I}_a$ respectively, and let $\mathcal{I}'_a$ denote the set $\mathcal{I}_a$ with F_a removed. By (3.3.19),

$$n2^{-a} \geq n - i[0] - \dots - i[f_a - 1] \geq n - i[0] - \dots - i[F_a - 1] \geq n2^{-a-1}, \quad (3.3.22)$$

and hence

$$\sum_{j \in \mathcal{I}'_a} i[j] \leq \sum_{j=f_a}^{F_a-1} i[j] \leq n2^{-a} - n2^{-a-1} = n2^{-a-1}. \quad (3.3.23)$$

Also, (3.3.15) and (3.3.19) imply that

$$\sum_{j \in \mathcal{I}'_a} i[j] \geq (|\mathcal{I}_a| - 1)(n2^{-a-1})^A. \quad (3.3.24)$$

Combining (3.3.23) and (3.3.24) yields (3.3.20), and the claim (3.3.18) follows.

To prepare for the application of Lemma 3.3.3, we let $k = k[u]$, $l = i[u]$, and $m = n - i[0] - \dots - i[u]$. (Recall that $k[u]$ and $i[u]$ are strictly positive.) Then (3.3.17) and (3.3.18) tell us that

$$h_n \leq \mu^{n-m} n^{2Cn^{1-A}+2} h_{m,k}^*. \quad (3.3.25)$$

If $m = 0$, then since $h_{0,k}^* = 1$ for every k , the result (3.3.3) follows from (3.3.25) by simply taking $A = d/(d+2)$. So for the remainder of the proof, we shall assume that m is strictly positive. By the definition of u , we know from (3.3.11), (3.3.12), and the bound $m + l \leq n$ that

$$\mu^l < (m+l)^2 b_{l,k} \leq n^2 b_{l,k} \quad (3.3.26)$$

and

$$l < (m+l)^A \leq n^A. \quad (3.3.27)$$

Combining (3.3.26) with Lemma 3.3.3 yields

$$n^{-2} \mu^l h_{m,k}^* \leq \mu^{m+l+dV} [D(m+l)]^{12m^B+3dV}, \quad (3.3.28)$$

which implies

$$h_{m,k}^* \leq \mu^{m+dV} n^2 [Dn]^{12n^B+3dV}. \quad (3.3.29)$$

Applying (3.3.29) to (3.3.25), we obtain

$$h_n \leq \mu^{n+dV} n^{2Cn^{1-A}+4} [Dn]^{12n^B+3dV}. \quad (3.3.30)$$

Now, $V = (m^{1-B}l)^{1/d} \leq n^{(1-B+A)/d}$ by (3.3.27), so we see that there is a constant Q , depending only on A , B , and d , such that

$$h_n \leq \mu^n \exp[Q(n^{(1-B+A)/d} + n^{1-A} + n^B) \log n] \quad (3.3.31)$$

for every $n \geq 2$. Finally, set $A = d/(d+2)$ and $B = 2/(d+2)$, so that $1-A = (1-B+A)/d = 2/(d+2)$ (these are the optimal choices for A and B). This proves (3.3.3), and Theorem 3.3.1 follows.

Proof of Lemma 3.3.3. Let β be an arbitrary l -step bridge starting at the origin and having span k . Let η be an arbitrary m -step half-space walk starting at the origin and having span at most k . Now let

$$\mathcal{Y} = \{y \in \mathbf{Z}^d : y_1 \geq \eta_1(m) \text{ and } \|y - \eta(m)\|_\infty \leq V\}.$$

Then $\mathcal{Y}$ is a half-cube containing $(\lfloor V \rfloor + 1)(2\lfloor V \rfloor + 1)^{d-1}$ points of $\mathbf{Z}^d$, and hence

$$|\mathcal{Y}| > V^d = m^{1-B}l. \quad (3.3.32)$$

For each y in $\mathbf{Z}^d$, let $J(y)$ denote the number of pairs (i, j) ($0 \leq i \leq m$, $0 \leq j \leq l$) such that $\eta(i) - \beta(j) = y$. The sum of $J(y)$ over all $y \in \mathbf{Z}^d$ equals $(m+1)(l+1)$, and so the average value of $J(y)$ over $\mathcal{Y}$ is less than or equal to

$$\frac{(m+1)(l+1)}{|\mathcal{Y}|} \leq \frac{4ml}{m^{1-B}l} = 4m^B.$$

Thus we know that there exists a point Y in $\mathcal{Y}$ such that

$$J(Y) \leq 4m^B. \quad (3.3.33)$$

Let $g = \|Y - \eta(m)\|_1$; since $Y \in \mathcal{Y}$ we know that

$$0 \leq g \leq dV. \quad (3.3.34)$$

Now define a walk ρ , not necessarily self-avoiding, which consists of η , followed by a walk of minimal length from $\eta(m)$ to Y , followed by β (translated to begin at Y). Thus ρ has exactly $m + g + l$ steps. Observe that

$$0 < \rho_1(i) \leq Y_1 + k = Y_1 + \beta_1(l) = \rho_1(m + g + l) \quad \text{for all } i = 0, \dots, m + g + l \quad (3.3.35)$$

(this is because (i) $\rho_1(i) = \eta_1(i) \in (0, k]$ whenever $0 < i \leq m$, (ii) $\eta_1(m) \leq \rho_1(i) \leq Y_1$ whenever $m \leq i \leq m+g$, and (iii) $\rho_1(i) = Y_1 + \beta_1(i-m-g) \in [Y_1, Y_1+k]$ whenever $m+g \leq i \leq m+g+l$). Let T be the number of self-intersections of ρ (if $\rho(i) = z$ for exactly n different values of i , then we count $n-1$ self-intersections). There are exactly $J(Y)$ intersections of the first $m+1$ sites of ρ with the last $l+1$ sites, and at most $g-1$ intersections of $(\rho(m+1), \dots, \rho(m+g-1))$ with the rest of ρ ; therefore

$$T \leq J(Y) + g - 1 \leq 4m^B + dV - 1 \quad (3.3.36)$$

by (3.3.33) and (3.3.34).

It follows from (3.3.35) that ρ can be obtained by taking a (self-avoiding) bridge of span Y_1+k and adjoining at most T self-avoiding polygons (including possibly “degenerate” two-step polygons). To understand this, think of traversing ρ one step at a time. When ρ first intersects itself, remove the segment of ρ between the two visits to the site where the self-intersection occurs. The removed segment is a self-avoiding polygon with a distinguished site (where the intersection occurs) and orientation (corresponding to the direction in which ρ traversed the polygon). Observe that some of these “polygons” may consist of only two steps, the direction of the second being the opposite of the first. Accordingly, we define the number of two-step polygons to be $q_2 = d$. Now continue to traverse ρ , and repeat this procedure: the next time that ρ visits a site that it has already visited (excluding visits that occurred on the removed segment), remove the resulting polygon, and so on. Let φ be the part of ρ that is never removed; then φ is a bridge of span Y_1+k . (We remark that this procedure is essentially the same as the “loop-erasing” of Section 10.2.) Let $r = |\varphi|$, let t be the number of polygons that have been removed by this procedure, and let a_i be the number of steps in the i -th polygon. Then $r + a_1 + \dots + a_t = |\rho|$.

Now, the number of different p -step walks ρ that can give rise to a particular choice of φ , t , and $a_1, \dots, a_t$ is at most

$$\prod_{j=1}^t (2pa_j q_{a_j}), \quad (3.3.37)$$

because there are q_{a_j} choices for the j -th polygon, exactly a_j different points on the j -th polygon where it could be attached to the walk (i.e. to φ or one of the first $j-1$ polygons), at most p places on the walk where the j -th polygon can be attached (in fact, at most $|\varphi| + a_1 + \dots + a_{j-1}$ places), and two possible directions that the polygon can be traversed. [If a_j equals 2 for some j , then it in fact would have sufficed to have used $2pq_2$ for the j -th term in (3.3.37).] For every even $n \geq 2$ we have $q_n \leq (d-1)\mu^n$ [by (3.2.5)

and (3.2.9) for $n \geq 4$; the case $n = 2$ is obvious since $\mu \geq d$], and so the product (3.3.37) can be bounded above by

$$[2(d-1)p^2]^t \mu^{a_1+\dots+a_t} = [2(d-1)p^2]^t \mu^{|\rho|-r}.$$

Therefore the number of possible walks ρ is at most

$$\sum_{p=m+l}^{m+l+dV} \sum_{r=1}^p b_r \sum_{t=0}^H \sum_{\substack{a_1, \dots, a_t : \\ a_1 + \dots + a_t = p-r}} [2(d-1)p^2]^t \mu^{p-r}, \quad (3.3.38)$$

where $H = 4m^B + dV - 1$ [recall (3.3.36)]. Using the fact that there are at least $b_{l,k} h_{n,k}^*$ possible walks ρ , we conclude from (3.3.38) and the bound $b_r \leq \mu^r$ [Equation (1.2.17)] that

$$\begin{aligned} b_{l,k} h_{n,k}^* &\leq \sum_{p=m+l}^{m+l+dV} \sum_{r=1}^p \sum_{t=0}^H \sum_{\substack{a_1, \dots, a_t : \\ a_1 + \dots + a_t = p-r}} [2(d-1)p^2]^H \mu^p \\ &\leq \sum_{p=m+l}^{m+l+dV} p(H+1) p^H [2(d-1)p^2]^H \mu^p \\ &\leq \mu^{m+l+dV} 4(m+l+dV)^3 [2(d-1)(m+l+dV)^3]^{4m^B+dV-1}. \end{aligned} \quad (3.3.39)$$

Finally, we obtain the inequality of the lemma from (3.3.39) and $V \leq (ml)^{1/2} \leq (m+l)/2$. $\square$

3.4 Notes

Section 3.1. An earlier paper [Hammersley (1961b)] proved that $c_N \leq \mu^N \exp[O(N^{(d-1)/d} \log N)]$ for $d \geq 2$. The proof was more complicated than the proof of Theorem 3.1.1, but the methods were similar; they were also closely related to the methods of Theorem 3.3.1.

The asymptotics of the number of partitions of N can also be used to study “spiral” walks, which are self-avoiding walks in $\mathbb{Z}^2$ that cannot turn to the left. Guttmann and Wormald (1984) proved that the number of N -step spiral walks is $\exp[2\pi(N/3)^{1/2}] N^{-7/4} [C + O(N^{-1/2})]$, where $C = 4 \cdot 3^{5/4} / \pi$.

Corollary 3.1.8 is due to Kesten (1963).

Section 3.2. Theorem 3.2.3 and Corollary 3.2.6 are due to Hammersley (1961a), using essentially the same proofs as we present. This paper also

contains a proof that $\mu_{\text{Polygon}} = \mu$, using very different methods from ours. A result similar to Theorem 3.2.4 appears in Equation (3.7) of Kesten (1963).

Dubins *et al.* (1988) proved the following result about self-avoiding polygons in $\mathbb{Z}^2$. Consider the set of all N -step polygons that have the origin as one of their sites. Then the probability that the point $(\frac{1}{2}, \frac{1}{2})$ lies in the inside region of a polygon chosen at random from this set equals $\frac{1}{2} - \frac{1}{N}$. (Here, the “inside region” is the bounded subset of $\mathbb{R}^2$ whose boundary is the simple closed curve determined by the polygon.) They conjecture that the analogous probability for any other point (a, b) of $\mathbb{R}^2$ (with a and b non-integer) should likewise increase to $1/2$ as $N \rightarrow \infty$, but nothing is known even for $(a, b) = (\frac{3}{2}, \frac{1}{2})$.

Chapter 4

Decay of the two-point function

4.1 Properties of the mass

In this section we shall develop some fundamental properties of the mass m , which we originally defined in Equation (1.3.15) as follows:

$$m(z) = \liminf_{n \rightarrow \infty} \frac{-\log G_z(0, (n, 0, \dots, 0))}{n}. \quad (4.1.1)$$

Thus the mass describes the exponential decay rate of the two-point function. We shall see that the “lim inf” appearing in (4.1.1) is in fact a limit for every $z > 0$ except perhaps $z = z_c$, and that the two-point function decays (to leading order) like $\exp[-m(z)|x|_z]$ for $z < z_c$, for some norm¹ $|\cdot|_z$. We shall also show that $m(z)$ is a reasonably nice function, strictly positive below the critical point z_c , identically $-\infty$ above z_c , and decreasing to 0 as z approaches z_c from the left. Finally we shall prove that $m(z_c) = 0$ if the “bubble diagram” $B(z) \equiv \sum_x G_z(0, x)^2$ (see Section 1.5) is finite at the critical point $z = z_c$ (as it is for $d \geq 5$; see Corollary 6.1.7). It is expected that $m(z_c) = 0$ in all dimensions [in fact, $G_{z_c}(0, x)$ is believed to decay as a power law; see (1.4.8)], but it remains an open problem to prove this for $d = 2, 3, 4$. In particular, it is not even known rigorously that $G_{z_c}(0, x)$ is finite for any $x \neq 0$ in low dimensions.

¹Alternatively, one can work with the Euclidean norm and a direction-dependent mass $m[v; z] \equiv m(z)|v|_z$ for vectors $v \in \mathbb{R}^d$ such that $|v| = 1$. Then $G_z(0, x)$ decays like $\exp[-m[v; z]|x|]$ where $v = x/|x|$.

The term “mass” comes from quantum field theory, where the exponential decay rate of the theory’s two-point function defines the physical mass of the particles in the theory. In statistical mechanics, the mass tending to 0 is equivalent to the correlation length $\xi(z) = 1/m(z)$ tending to ∞ . In the context of the Ising model and the other N -vector models, for example, this says that spins are becoming correlated on larger and larger length scales. In a spin model, the divergence of the correlation length as $z \nearrow z_c$ is the precursor of the long range order (and spontaneous magnetization) that will occur for $z > z_c$. The self-avoiding walk corresponds to the $N = 0$ case, where the concept of long-range order does not really apply, but we are still interested in whether the mass tends to 0 by analogy to spin systems. In addition, it is believed (and known for $d \geq 5$) that the mass for the self-avoiding walk goes to zero as a power law

$$m(z) \sim \text{const.} (z_c - z)^{\nu} \quad \text{as } z \nearrow z_c, \quad (4.1.2)$$

with $\nu = \nu$ (recall Section 1.3). Proving that $m(z) \searrow 0$ as $z \nearrow z_c$ is a first step towards proving (4.1.2).

Many results of this section extend immediately to self-avoiding walks $(\omega(0), \dots, \omega(N))$ whose steps $\omega(i) - \omega(i-1)$ all lie in a finite subset Ω of $\mathbb{Z}^d$ which is invariant under all symmetries of $\mathbb{Z}^d$. Besides the usual nearest-neighbour model ($\Omega = \{x : \|x\|_1 = 1\}$), in Chapter 6 we shall also be interested in the “spread-out” models which have $\Omega = \{x : 0 < \|x\|_\infty \leq L\}$ for some (large) integer L . In particular, everything in this section up to and including Theorem 4.1.6, as well as Theorem 4.1.18, hold for general symmetric Ω with only trivial changes in the proofs.

Our first proposition establishes some elementary properties.

Proposition 4.1.1 (a) $m(z)$ is a concave function of $\log z$ for $z > 0$.
 (b) On the interval $(0, z_c)$, $m(z)$ is a nonincreasing, finite, strictly positive, and continuous function of z .
 (c) If $z > z_c$, then $G_z(0, x) = +\infty$ for every $x \neq 0$, and hence $m(z) = -\infty$.

We delay the proof of this proposition just long enough to present the following lemma:

Lemma 4.1.2 Let $\{a_n\}_{n \geq 0}$ be a sequence of nonnegative numbers. Then $-\log(\sum_n a_n e^{n\beta})$ is a concave function of β .

Proof. This is a consequence of Hölder’s inequality. For λ between 0 and 1,

$$\sum_n a_n e^{n[\lambda\beta_1 + (1-\lambda)\beta_2]} \leq \left(\sum_n a_n e^{n\beta_1}\right)^\lambda \left(\sum_n a_n e^{n\beta_2}\right)^{1-\lambda}.$$

The lemma follows upon taking $-\log$ of both sides. $\square$

Proof of Proposition 4.1.1. (a) Lemma 4.1.2 shows that $-\log G_z(0, x)$ is a concave function of $\log z$ for any x . Since the $\liminf$ of a sequence of concave functions is concave, the result follows.

(b) Since $G_z(0, x)$ is nondecreasing in z , it is apparent from (4.1.1) that $m(z)$ is nonincreasing. We already saw in Section 1.3 that $0 < m(z) < +\infty$ whenever $0 < z < z_c$ [recall (1.3.14) and (1.3.16)]. Continuity follows from the concavity and finiteness of $m(z)$ on the open interval $(0, z_c)$.

(c) This is an immediate consequence of Corollary 3.2.6. $\square$

Next, we shall show how to replace the “ $\liminf$ ” in (4.1.1) by a limit, obtaining as a by-product an explicit bound on the two-point function in terms of the mass and the bubble diagram $B(z) = \sum_x G_z(0, x)^2$. We will use the notation $(n, 0)$ to denote the point $(n, 0, \dots, 0) \in \mathbb{Z}^d$; this notation will be generalized in Definition 4.1.7 below.

Theorem 4.1.3 (a) If $0 < z < z_c$, then

$$\lim_{n \rightarrow \infty} \frac{-\log G_z(0, (n, 0))}{n} = m(z) = \inf_{n \geq 1} \frac{-\log[G_z(0, (n, 0))/B(z)]}{n}; \quad (4.1.3)$$

in particular, the limit exists and satisfies

$$G_z(0, (n, 0)) \leq B(z)e^{-m(z)n} \quad \text{for every } n \geq 1. \quad (4.1.4)$$

(b) If $B(z_c)$ is finite, then (4.1.3) and (4.1.4) also hold for $z = z_c$, and $m(z)$ is left-continuous at z_c .

The proof depends on subadditivity and the following lemma.

Lemma 4.1.4 For any $z > 0$, and any x and y in $\mathbb{Z}^d$,

$$G_z(0, x)G_z(x, y) \leq B(z)G_z(0, y). \quad (4.1.5)$$

Proof. For each nonnegative integer N , let $\mathcal{S}_N$ denote the set of all ordered pairs of self-avoiding walks (ω_A, ω_B) such that: ω_A starts at 0 and ends at x ; ω_B starts at x and ends at y ; and $|\omega_A| + |\omega_B| = N$. Also, let $\mathcal{T}_N$ denote the set of all ordered triples of self-avoiding walks $(\omega_C, \omega_D, \omega_E)$ such that: ω_C starts at 0 and ends at y ; ω_D and ω_E both start at x and end at the same (arbitrary) point; and $|\omega_C| + |\omega_D| + |\omega_E| = N$. To prove the lemma, it suffices to show that there is a one-to-one mapping from $\mathcal{S}_N$ into $\mathcal{T}_N$, for this would imply an inequality between their respective generating functions, which is precisely the inequality that we want.

Let (ω_A, ω_B) be a member of S_N . Let I be the smallest value of i such that $\omega_A(i)$ is a point of ω_B . Let $u = \omega_A(I)$. Let ω_C be the walk which follows ω_A from 0 to u and then follows ω_B from u to y ; this is self-avoiding by our choice of u . Let ω_D (respectively, ω_E) be the part of ω_A (respectively, ω_B) between x and u . Then $(\omega_C, \omega_D, \omega_E)$ is in T_N . This mapping is clearly one-to-one, so the lemma is proven. $\square$

Proof of Theorem 4.1.3. Fix z with $B(z) < \infty$, so in particular we may choose any z in $(0, z_c)$ since $B(z) \leq \chi(z)^2$. For each integer $n \geq 1$, define $h_n(z) = -\log[G_z(0, (n, 0))/B(z)]$. Taking $x = (n, 0)$ and $y = (m+n, 0)$ in (4.1.5) and dividing by $B(z)^2$ shows that the sequence $\{h_n(z) : n \geq 1\}$ is subadditive. Therefore Lemma 1.2.2 implies that

$$\lim_{n \rightarrow \infty} \frac{h_n(z)}{n} = \inf_{n \geq 1} \frac{h_n(z)}{n}, \quad (4.1.6)$$

which proves (4.1.3) in both cases (a) and (b). The bound (4.1.4) follows immediately from (4.1.3).

It only remains to prove that $m(z)$ is left-continuous at z_c if $B(z_c)$ is finite. Since $m(z)$ is nonincreasing, it suffices to show that

$$\limsup_{z \nearrow z_c} m(z) \leq m(z_c). \quad (4.1.7)$$

Assume $B(z_c)$ is finite. This implies that $G_{z_c}(0, x)$ is finite (for every x); hence, since $B(z)$ and $G_z(0, x)$ are power series with nonnegative coefficients, they must be continuous on $(0, z_c]$. Therefore $h_n(z)$ is continuous on $(0, z_c]$ for every n . Together with the fact that $m(z) \leq h_n(z)/n$ for all z in $(0, z_c)$ [by (4.1.3)], this implies that

$$\limsup_{z \nearrow z_c} m(z) \leq \limsup_{z \nearrow z_c} \frac{h_n(z)}{n} = \frac{h_n(z_c)}{n}. \quad (4.1.8)$$

Finally, we have seen that $m(z_c) = \inf_{n \geq 1} h_n(z_c)/n$ when $B(z_c)$ is finite, so taking the inf over $n \geq 1$ in (4.1.8) yields (4.1.7), which completes the proof. $\square$

We now turn our attention to the task of showing that the mass goes to 0 as z approaches z_c from the left. This turns out to be relatively easy if the bubble condition $B(z_c) < \infty$ holds. However, it is expected that $B(z_c)$ is infinite in 2, 3, and 4 dimensions, so we will have to work harder there. But first we shall take care of the high-dimensional case.

Lemma 4.1.5 For any $z > 0$ and any x in $\mathbb{Z}^d$,

$$G_z(0, x) \leq B(z)^{1/2} e^{-m(z)\|x\|_\infty}. \quad (4.1.9)$$

Proof. Given x in $\mathbb{Z}^d$, let $K = \|x\|_\infty$ and let i be a coordinate such that $|x_i| = K$. Let u be the vector whose i -th coordinate is x_i and whose j -th coordinate, for every $j \neq i$, is $-x_j$. Then $x + u = \pm 2Ke^{(i)}$, where $e^{(i)}$ is the unit vector whose i -th coordinate is 1. By Lemma 4.1.4 and symmetry considerations, we have

$$G_z(0, x)G_z(x, x+u) \leq B(z)G_z(0, x+u) = B(z)G_z(0, (2K, 0)). \quad (4.1.10)$$

Applying symmetry and (4.1.4), we obtain $G_z(0, x)^2 \leq B(z) \exp[-m(z)2K]$, and (4.1.9) follows. $\square$

Theorem 4.1.6 If $B(z_c) < \infty$, then $\lim_{z \nearrow z_c} m(z) = m(z_c) = 0$.

Proof. The first equality was proven in Theorem 4.1.3(b). Since $m(z) > 0$ for $z < z_c$ [by Proposition 4.1.1(b)], we see that $m(z_c) \geq 0$. Finally, if $m(z_c)$ were strictly positive, then Lemma 4.1.5 would imply that the critical two-point function decays exponentially, which would contradict the fact that the susceptibility is infinite at z_c [recall (1.3.6)]. Therefore $m(z_c)$ must equal 0. $\square$

For the rest of this chapter, we will not assume that the bubble condition holds. We will not be able to prove that the mass is 0 at z_c , but we will show that the mass decreases to 0 as z approaches z_c from the left. Some new ideas will be needed to accomplish this. In a nutshell, we would like a subadditivity relation in the spirit of Lemma 4.1.4 that holds nontrivially at the critical point. As in Section 3.1, we shall use bridges to get superadditivity relations instead [e.g. (4.1.13) and (4.1.14) below]. We first define generating functions and masses for classes of bridges, prove properties about these, and then show that these masses are the same as the one defined by (4.1.1). As the reader will discover, it is often easier to work with bridges than with general self-avoiding walks.

Definition 4.1.7 Let $y = (y_1, \dots, y_{d-1})$ be a point of $\mathbb{Z}^{d-1}$, and let L be a nonnegative integer. Then (L, y) denotes the point $(L, y_1, \dots, y_{d-1})$ in $\mathbb{Z}^d$, and $b_{N,L}(y)$ denotes the number of N -step bridges ω with $\omega(0) = 0$ and $\omega(N) = (L, y)$. Recalling Definition 3.1.3, we see that

$$b_{N,L} = \sum_{y \in \mathbb{Z}^{d-1}} b_{N,L}(y).$$

For each real $z > 0$, we define the generating functions

$$B_z(L, y) = \sum_{N=0}^{\infty} b_{N,L}(y) z^N$$

(the "point-to-point" bridge generating function) and

$$B_z(L) = \sum_{N=0}^{\infty} b_{N,L} z^N$$

(the "point-to-plane" bridge generating function). Observe that

$$B_z(L) = \sum_{y \in \mathbb{Z}^{d-1}} B_z(L, y).$$

Remark. To be fully consistent with our notation for the two-point function, we should be writing $B_z(0, (L, y))$ instead of $B_z(L, y)$. However, the shorter notation should not cause any confusion.

Proposition 4.1.8 For every real $z > 0$, the limits

$$M(z) = \lim_{L \rightarrow \infty} \frac{-\log B_z(L, 0)}{L} \text{ and } \overline{M}(z) = \lim_{L \rightarrow \infty} \frac{-\log B_z(L)}{L} \quad (4.1.11)$$

exist in $[-\infty, +\infty)$, and satisfy

$$B_z(L, 0) \leq e^{-LM(z)} \text{ and } B_z(L) \leq e^{-L\overline{M}(z)} \quad (4.1.12)$$

for every integer $L \geq 1$. Also, $M(z)$ and $\overline{M}(z)$ are nonincreasing functions of z , and they satisfy the obvious inequalities $M(z) \geq \overline{M}(z)$ and $M(z) \geq m(z)$.

Proof. Let L_1 and L_2 be nonnegative integers. The concatenation of a bridge of span L_1 with a bridge of span L_2 is a bridge of span $L_1 + L_2$, and the result uniquely determines the original pair. Thus it is apparent that

$$\sum_{n=0}^N b_{n,L_1} b_{N-n,L_2} \leq b_{N,L_1+L_2}$$

and

$$\sum_{n=0}^N b_{n,L_1}(0) b_{N-n,L_2}(0) \leq b_{N,L_1+L_2}(0)$$

for every nonnegative integer N , so

$$B_z(L_1) B_z(L_2) \leq B_z(L_1 + L_2) \quad (4.1.13)$$

and

$$B_z(L_1, 0) B_z(L_2, 0) \leq B_z(L_1 + L_2, 0) \quad (4.1.14)$$

for all $z > 0$. Therefore, by Lemma 1.2.2, the limits in (4.1.11) exist and satisfy (4.1.12) for all $L \geq 1$. $\square$

We will now develop some properties of the mass M , and eventually we will show that M and $\overline{M}$ are identical. We begin with an analogue of Proposition 4.1.1.

Proposition 4.1.9 (a) $M(z) \leq -\log z$ for all $z > 0$.

(b) $M(z) \geq -\log(\mu z)$ for $0 < z < z_c$.

(c) M is a concave function of $\log z$ ($z > 0$).

In particular, M is finite and continuous on $(0, z_c)$.

Proof. (a) For every L , $b_{L,L}(0) = 1$, so $B_z(L, 0) \geq z^L$. The result follows now from (4.1.12).

(b) Fix z in $(0, z_c)$. Since $b_N \leq \mu^N$ by (1.2.17),

$$B_z(L, 0) \leq \sum_{N=L}^{\infty} \mu^N z^N = \frac{(\mu z)^L}{1 - \mu z}$$

for every $L > 0$. The result follows from (4.1.11).

(c) This follows by applying Lemma 4.1.2 with $a_n = b_{n,L}(0)$, dividing by L , and letting $L \rightarrow \infty$. $\square$

We now describe "truncated" generating functions for bridges that are confined to a tube centred along the x_1 -axis. The lemma which follows shows that the corresponding truncated mass converges to the mass $M(z)$ as the radius of the tube tends to infinity.

Definition 4.1.10 For all positive integers N , L , and T , and for all points y in $\mathbb{Z}^{d-1}$, let $b_{N,L}^T(y)$ be the number of N -step bridges ω having $\omega(0) = 0$, $\omega(N) = (L, y)$, and $|\omega_i(k)| \leq T$ for every $i = 2, \dots, d$ and $k = 0, \dots, N$. For real $z > 0$, let

$$B_z^T(L, y) = \sum_N b_{N,L}^T(y) z^N.$$

Observe that the Monotone Convergence Theorem implies that for every $z > 0$

$$\lim_{T \rightarrow \infty} B_z^T(L, y) = B_z(L, y) \text{ for all } L \text{ and } y. \quad (4.1.15)$$

Also, by the usual concatenation argument, for every T we have

$$B_z^T(L_1, 0) B_z^T(L_2, 0) \leq B_z^T(L_1 + L_2, 0) \text{ for all } L_1 \text{ and } L_2, \quad (4.1.16)$$

which implies, by Lemma 1.2.2, that we can define the "truncated masses"

$$M^T(z) = \lim_{L \rightarrow \infty} \frac{-\log B_z^T(L, 0)}{L} = \inf_{L \geq 1} \frac{-\log B_z^T(L, 0)}{L}. \quad (4.1.17)$$

Lemma 4.1.11 Let T and L be positive integers, and let z be a positive real number.

(a) Let $M_L^T(z) = -(\log B_z^T(L, 0))/L$. Then

$$-\frac{(2T+1)^{d-1}}{z} \leq \frac{d}{dz} M_L^T(z) \leq -\frac{1}{z}.$$

(b) For $z_2 > z_1 > 0$:

$$-(2T+1)^{d-1} \log(z_2/z_1) \leq M^T(z_2) - M^T(z_1) \leq -\log(z_2/z_1).$$

In particular, M^T is a continuous decreasing function of z .

(c) $\lim_{T \rightarrow \infty} M^T(z) = \inf_{T \geq 1} M^T(z) = M(z)$.

(d) M is left-continuous; i.e., $\lim_{u \nearrow z} M(u) = M(z)$ for all $z > 0$.

In (c) and (d), the limits may be $-\infty$.

Remark. Part (d) of this lemma is mainly of interest at the critical point, since $M(z)$ is already known to be continuous on $(0, z_c)$ by Proposition 4.1.9.

Proof of Lemma 4.1.11. (a) First observe that $B_z^T(L, 0)$ is a polynomial with positive coefficients, and so $M^T(z)$ is differentiable at every $z > 0$:

$$\frac{d}{dz} M_L^T(z) = -\frac{\sum_N N b_{N,L}^T(0) z^{N-1}}{L \sum_N b_{N,L}^T(0) z^N}.$$

Since $b_{N,L}^T(0)$ is nonzero only if N is between L and $L(2T+1)^{d-1}$, the result follows.

(b) The result follows upon integrating the inequalities of part (a) from z_1 to z_2 and then letting $L \rightarrow \infty$ [using (4.1.17)].

(c) Since $M^T(z)$ is decreasing in T , it suffices to show that $\inf_{T \geq 1} M^T(z) = M(z)$. By subadditivity (recall Proposition 4.1.8) and (4.1.15),

$$M(z) = \inf_{L \geq 1} \frac{-\log B_z(L, 0)}{L} = \inf_{L \geq 1} \inf_{T \geq 1} \frac{-\log B_z^T(L, 0)}{L}.$$

The result now follows by interchanging the order of the infs in the last expression and using (4.1.17).

(d) This follows from parts (b) and (c), together with the general fact that the inf of a sequence of continuous decreasing functions is left-continuous. $\square$

Lemma 4.1.12 $M(z) = \overline{M}(z)$ for all $z > 0$.

Proof. Since $M(z) \geq \overline{M}(z)$, it suffices to prove the reverse inequality. To do this, we use a slightly different form of truncation. Let $\overline{B}_z^T(L, y)$ denote the generating function of the collection of bridges ω from 0 to (L, y) with the property that

$$|\omega_i(j) - \omega_i(k)| \leq T \text{ for } 2 \leq i \leq d \text{ whenever } \omega_1(j) = \omega_1(k).$$

Let $\overline{B}_z^T(L) = \sum_y \overline{B}_z^T(L, y)$. The analogue of (4.1.16) holds for $\overline{B}_z^T(L)$, and hence the mass $\overline{M}^T(z)$ of $\overline{B}_z^T(L)$ exists and satisfies the analogue of (4.1.17). The same arguments as in the proof of Lemma 4.1.11(c) show that $\lim_{T \rightarrow \infty} \overline{M}^T(z) = \overline{M}(z)$. Observe that $\overline{B}_z^T(L, y) = 0$ if $\|y\|_\infty > LT$.

For any $L > 0$ and any y in $\mathbb{Z}^{d-1}$, we can get a bridge from 0 to $(2L+1, 0)$ by the concatenation of a bridge from 0 to (L, y) , a single step from (L, y) to $(L+1, y)$, and another bridge from 0 to (L, y) that has been reflected through the hyperplane $x_1 = L + \frac{1}{2}$. This construction and the Schwarz inequality show that, for any $T > 0$,

$$\begin{aligned} B_z(2L+1, 0) &\geq \sum_{y \in \mathbb{Z}^{d-1}: \|y\|_\infty \leq LT} z(\overline{B}_z^T(L, y))^2 \\ &\geq \frac{z(\sum_{y: \|y\|_\infty \leq LT} \overline{B}_z^T(L, y))^2}{\sum_{y: \|y\|_\infty \leq LT} 1^2} \\ &= \frac{z(\overline{B}_z^T(L))^2}{(2LT+1)^{d-1}}. \end{aligned}$$

This implies that for every fixed T , $M(z) \leq \overline{M}^T(z)$. The lemma follows. $\square$

Theorem 4.1.13 $\lim_{z \nearrow z_c} M(z) = M(z_c) = 0$.

Proof. First, Lemma 4.1.11(d) says that M is left-continuous. By Proposition 4.1.9(b), we know that $M(z) > 0$ whenever $0 < z < z_c$, and so $M(z_c) \geq 0$. Next, Lemma 4.1.12 tells us that $M(z_c) = \overline{M}(z_c)$, so it only remains to prove that $\overline{M}(z_c) \leq 0$. But if it were true that $\overline{M}(z_c) > 0$, then it would follow from (4.1.12) that

$$B_{z_c} = \sum_{L=0}^{\infty} B_{z_c}(L) \leq \sum_{L=0}^{\infty} e^{-L\overline{M}(z_c)} < +\infty,$$

which contradicts Corollary 3.1.8. This completes the proof. $\square$

We are finally ready to prove that the various masses are identical below the critical point.

Theorem 4.1.14 For all z in $(0, z_c)$, $m(z) = M(z) = \overline{M}(z)$.

Proof. Fix z in $(0, z_c)$. Recall $M(z) = \overline{M}(z)$ by Lemma 4.1.12. Since $G_z(0, (n, 0)) \geq B_z(n, 0)$, it is clear that $m(z) \leq M(z)$.

Any self-avoiding walk from 0 to $(n, 0)$ can be decomposed into three parts: cut the walk at the last time that it visits the hyperplane $x_1 = 0$ and at the first time after this that it visits the hyperplane $x_1 = n$. The middle piece is a bridge of span n ; the other two pieces are self-avoiding walks (possibly having length 0). This decomposition shows that

$$G_z(0, (n, 0)) \leq (\chi(z))^2 B_z(n) \leq (\chi(z))^2 e^{-M(z)n}, \quad (4.1.18)$$

where we have also used (4.1.12) and $\overline{M}(z) = M(z)$. Since $z < z_c$, $\chi(z)$ is finite, so $m(z) \geq M(z)$. The theorem follows. $\square$

Corollary 4.1.15 $\lim_{z \nearrow z_c} m(z) = 0$.

Proof. This is an immediate consequence of Theorems 4.1.14 and 4.1.13. $\square$

Corollary 4.1.16 The mass $m(z)$ is strictly decreasing on $(0, z_c)$, and $\lim_{z \searrow 0} m(z) = +\infty$.

Proof. By Lemma 4.1.11(b,c), the function $M(z) + \log z$ is nonincreasing on $(0, +\infty)$. Since $M(z)$ is finite and equals $m(z)$ on $(0, z_c)$, the corollary follows. $\square$

Corollary 4.1.17 Define

$$G_z(L) = \sum_{y \in \mathbb{Z}^{d-1}} G_z(0, (L, y)), \quad (4.1.19)$$

the generating function of all self-avoiding walks from the origin to the hyperplane $x_1 = L$. Then

$$\lim_{L \rightarrow \infty} \frac{-\log G_z(L)}{L} = m(z) \text{ for all } z \text{ in } (0, z_c). \quad (4.1.20)$$

In fact,

$$e^{-m(z)L} \leq G_z(L) \leq \chi(z)^2 e^{-m(z)L} \text{ for all } L \geq 1. \quad (4.1.21)$$

Proof. Equation (4.1.20) follows from

$$B_z(L) \leq G_z(L) \leq \chi(z)^2 B_z(L), \quad (4.1.22)$$

which may be obtained by the same argument that gave (4.1.18). The upper bound of (4.1.21) follows from (4.1.22) and $B_z(L) \leq e^{-m(z)L}$. The lower bound of (4.1.21) follows from the subadditivity relation (1.2.5), since $G_z(L_1)G_z(L_2) \geq G_z(L_1 + L_2)$ whenever L_1 and L_2 are positive integers. $\square$

We now show that Theorem 4.1.3 can be generalized so that x can tend to infinity in any direction. We remark that the proof of the following theorem relies only on material from the first part of this section (prior to Definition 4.1.7). Recall that the bubble diagram $B(z) = \sum_x G_z(0, x)^2$ is finite for $0 < z < z_c$ (since $B(z) \leq \chi(z)^2$).

Theorem 4.1.18 For any $0 < z < z_c$, there exists a norm $|\cdot|_z$ on $\mathbb{R}^d$, satisfying $\|u\|_\infty \leq |u|_z \leq \|u\|_1$ for every u in $\mathbb{R}^d$, such that

$$\lim_{|x|_z \rightarrow \infty} \frac{-\log G_z(0, x)}{|x|_z} = m(z) \quad (4.1.23)$$

and

$$G_z(0, x) \leq B(z)e^{-m(z)|x|_z} \text{ for every } x \text{ in } \mathbb{Z}^d. \quad (4.1.24)$$

Proof. Fix z in $(0, z_c)$. For each v in $\mathbb{Z}^d$, Lemma 4.1.4 tells us that

$$\frac{G_z(0, jv)}{B(z)} \frac{G_z(0, kv)}{B(z)} \leq \frac{G_z(0, (j+k)v)}{B(z)}$$

for all nonnegative integers j and k . Therefore Lemma 1.2.2 implies that the limit

$$\lim_{n \rightarrow \infty} \frac{-\log G_z(0, nv)}{n}$$

exists, and, if we denote this limit by $m[v; z]$, that

$$G_z(0, nv) \leq B(z)e^{-nm[v; z]} \text{ for every } n \geq 1. \quad (4.1.25)$$

We have $m[0; z] = 0$ since $G_z(0, 0) = 1$, but for every nonzero v in $\mathbb{Z}^d$ we have $0 < m[v; z] \leq \|v\|_1 \log z$ (the lower bound follows from (1.3.14), while

$$G_z(0, v) \geq z^{\|v\|_1} \quad (4.1.26)$$

gives the upper bound). Also, it follows immediately that

$$m[kv; z] = |k|m[v; z] \quad (4.1.27)$$

for every integer k and every v in $\mathbf{Z}^d$, and (from Lemma 4.1.4) that

$$m[u + v; z] \leq m[u; z] + m[v; z]$$

for every u and v in $\mathbf{Z}^d$.

Define $|v|_z$ by

$$|v|_z = \frac{m[v; z]}{m(z)}$$

for v in $\mathbf{Z}^d$, and use (4.1.27) to extend the definition to v in $\mathbf{Q}^d$ (where $\mathbf{Q}$ denotes the rational numbers). Observe that $|u|_z = 1$ if u is one of the $2d$ unit vectors of $\mathbf{Z}^d$ (by Theorem 4.1.3), that $|kv|_z = |k||v|_z$ for every rational k and every v in $\mathbf{Q}^d$, and that $|u + v|_z \leq |u|_z + |v|_z$ for every u and v in $\mathbf{Q}^d$; in particular, then, we have $|v|_z \leq \|v\|_1$ for every v in $\mathbf{Q}^d$. Consequently, we obtain

$$||u|_z - |v|_z| \leq |u - v|_z \leq \|u - v\|_1;$$

thus, $|\cdot|_z$ is uniformly continuous on $\mathbf{Q}^d$, and so it extends to a continuous function on all of $\mathbf{R}^d$, which will be a norm on $\mathbf{R}^d$. Next, Lemma 4.1.5 shows that $m[v; z] \geq m(z)|v|_\infty$ for every v in $\mathbf{Z}^d$. From this it follows that $\|v\|_\infty \leq |v|_z$ on $\mathbf{Z}^d$, and hence on all of $\mathbf{R}^d$.

From (4.1.25) and the definition of $|\cdot|_z$, we obtain

$$G_z(0, v) \leq B(z) \exp(-|v|_z m(z)) \quad \text{for every } v \text{ in } \mathbf{Z}^d, \quad (4.1.28)$$

which is (4.1.24). It follows from this that if (4.1.23) is false then there exists a sequence $\{x_n\}$ of points in $\mathbf{Z}^d$, tending to infinity in norm, such that

$$\lim_{n \rightarrow \infty} \frac{-\log G_z(0, x_n)}{|x_n|_z} > m(z). \quad (4.1.29)$$

By choosing a subsequence if necessary, we can also assume that there exists a t in $\mathbf{R}^d$ such that $x_n/|x_n|_z$ converges to t . Let $\epsilon > 0$. Since $|t|_z = 1$, we can choose a v in $\mathbf{Z}^d$ and a positive integer J such that

$$\|t - J^{-1}v\|_1 \leq \epsilon \quad \text{and} \quad |J^{-1}v|_z \leq 1 + \epsilon.$$

Next, choose a sequence of integers $k(n)$ such that $k(n)/|x_n|_z$ tends to J^{-1} . When n is large, x_n is approximately $k(n)v$; this will lead to a contradiction of (4.1.29), as follows.

Lemma 4.1.4 tells us that

$$G_z(0, k(n)v)G_z(k(n)v, x_n) \leq B(z)G_z(0, x_n).$$

Now take logs of this inequality, divide by $-|x_n|_z$, and let n tend to infinity. Using

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{-\log G_z(0, k(n)v)}{|x_n|_z} &= J^{-1}m[v; z] \\ &= |J^{-1}v|_z m(z) \\ &\leq (1 + \epsilon)m(z) \end{aligned}$$

and [with the help of (4.1.26)]

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{-\log G_z(k(n)v, x_n)}{|x_n|_z} &\leq -\lim_{n \rightarrow \infty} \frac{\|k(n)v - x_n\|_1}{|x_n|_z} \log z \\ &= -\|J^{-1}v - t\|_1 \log z \\ &\leq -\epsilon \log z, \end{aligned}$$

we conclude that

$$\lim_{n \rightarrow \infty} \frac{-\log G_z(0, x_n)}{|x_n|_z} \leq (1 + \epsilon)m(z) - \epsilon \log z.$$

Since ϵ can be made arbitrarily close to 0, this contradicts our choice of $\{x_n\}$ and completes the proof of the theorem. $\square$

4.2 Bridges and renewal theory

The main goal of the rest of this chapter is to give a more refined analysis of the asymptotics of $G_z(0, (L, 0, \dots, 0))$ for large L when $z < z_c$. Most of the work can be done by focusing first on bridges and their two-point functions. The present section will give the asymptotic behaviour of the bridge generating functions. Specifically, we shall prove in Theorem 4.2.5 that $B_z(L)$ exhibits pure exponential decay, and in Theorem 4.2.6 that $B_z(L, y)$ exhibits exponential decay with a Gaussian power law correction, also known as *Ornstein-Zernike decay*. We shall also prove that the mass $m(z)$ is a real analytic function of z in the interval $(0, z_c)$. To obtain these results and develop some intuition about why they are true, we shall set up a correspondence between generating functions and certain probabilistic quantities in renewal theory.

We begin by discussing the decomposition of bridges. Given an N -step bridge ω , suppose that i satisfies $0 < i \leq N$ and

$$\omega_1(j) \leq \omega_1(i) < \omega_1(k) \quad \text{for all } j = 0, \dots, i \text{ and } k = i + 1, \dots, N. \quad (4.2.1)$$

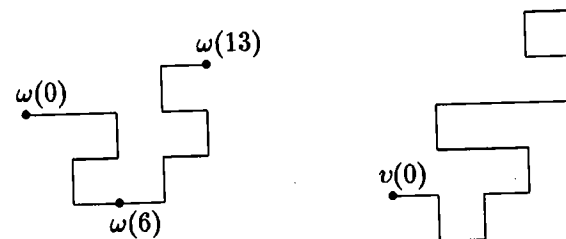


Figure 4.1: The bridge ω on the left can be decomposed into two smaller bridges: a 6-step bridge followed by a 7-step bridge. The bridge v on the right is irreducible.

Then ω can be decomposed into two smaller bridges, $(\omega(0), \dots, \omega(i))$ and $(\omega(i), \dots, \omega(N))$. (See Figure 4.1.) Observe that (4.2.1) always holds for $i = N$; in this case ω is trivially decomposed into ω and the 0-step bridge.

Definition 4.2.1 We say that an N -step bridge is irreducible if the only i ($0 < i \leq N$) for which (4.2.1) holds is $i = N$. Let λ_N denote the number of irreducible N -step bridges, and let

$$\Lambda_z = \sum_{N=1}^{\infty} \lambda_N z^N$$

be the corresponding generating function.

Observe that if ω is an irreducible N -step bridge, then for each $a = 1, \dots, \omega_1(N)$, there exist at least three distinct values of i such that $\omega_1(i) = a$.

Given an N -step bridge ω (with $N > 0$), let s be the smallest index i for which (4.2.1) holds. Then $(\omega(0), \dots, \omega(s))$ is an irreducible bridge and $(\omega(s), \dots, \omega(N))$ is a bridge. It is thus straightforward to see that

$$b_N = \sum_{s=1}^N \lambda_s b_{N-s} + \delta_{N,0} \quad (4.2.2)$$

for every $N \geq 0$. From this equation we immediately obtain

$$B_z = \frac{1}{1 - \Lambda_z} \text{ for all complex } z \text{ with } |z| < z_c, \quad (4.2.3)$$

where B_z is the generating function for the number of bridges (recall Definition 3.1.7). For $0 < z < z_c$, B_z is finite, and so $\Lambda_z < 1$; therefore $\Lambda_{z_c} \leq 1$

by the monotone convergence theorem. Also, B_z diverges at z_c (Corollary 3.1.8), so in fact $\Lambda_{z_c} = 1$; that is,

$$\sum_{k=1}^{\infty} \frac{\lambda_k}{\mu^k} = 1. \quad (4.2.4)$$

(It turns out that z_c is the radius of convergence of Λ_z ; see Corollary 4.4.5.)

Equation (4.2.2) may now be transformed into a probabilistic renewal equation as follows. Let $p_k = \lambda_k \mu^{-k}$ (for $k \geq 1$) and $a_N = b_N \mu^{-N}$ (for $N \geq 0$). Observe that $a_0 = 1$, $a_k \leq 1$ by (1.2.17), and $\sum_k p_k = 1$ by (4.2.4). Multiplying (4.2.2) by μ^{-N} yields, for $N \geq 1$,

$$a_N = \sum_{k=1}^N p_k a_{N-k}. \quad (4.2.5)$$

To interpret this probabilistically, suppose that we have an independent sequence of random variables $X_1, X_2, \dots$ with common distribution $\Pr\{X = k\} = p_k$. Then

$$a_N = \Pr\{X_1 + \dots + X_k = N \text{ for some } k \geq 0\}; \quad (4.2.6)$$

i.e., a_N is the probability that there is a “renewal” at “time” N . (To understand this terminology, one can think of the X_i ’s as representing the lifetimes of light bulbs, where a new bulb immediately replaces one that burns out. Then a_N is the probability that a new bulb is installed on day N —i.e., that the system is renewed on day N .) Taking (4.2.6) as the definition of the sequence a_N , it is easy to verify (4.2.5). This probabilistic interpretation will be exploited later.

Equation (4.2.5) is a *discrete renewal equation*. The main theorem about these equations is the Renewal Theorem, which we state in the following form.

Theorem 4.2.2 Assume that $\{f_n : n \geq 1\}$ and $\{g_n : n \geq 0\}$ are nonnegative sequences, and let

$$f = \sum_{n=1}^{\infty} f_n \quad \text{and} \quad g = \sum_{n=0}^{\infty} g_n$$

denote their sums. Assume that $0 < g < +\infty$ and that $f_1 > 0$. Define the new sequence $v_0, v_1, \dots$ by

$$\begin{aligned} v_0 &= g_0 \\ v_n &= g_n + f_1 v_{n-1} + f_2 v_{n-2} + \dots + f_n v_0, \text{ for all } n \geq 1. \end{aligned}$$

- (a) If $f < 1$, then $\lim_{n \rightarrow \infty} v_n = 0$ and $\sum_{n=0}^{\infty} v_n = g/(1-f)$.
 (b) If $f = 1$, then

$$\lim_{n \rightarrow \infty} v_n = \frac{g}{\sum_{k=1}^{\infty} k f_k}$$

(the limit is 0 if the sum in the denominator diverges). Also, $\sum_n v_n$ diverges.

- (c) If $f > 1$, then $\limsup_{n \rightarrow \infty} v_n^{1/n} > 1$.

In the usual (more general) statements of the Renewal Theorem, the condition $f_1 > 0$ is replaced by the condition that the greatest common divisor of $\{n : f_n > 0\}$ is one. Also, there are more complete results available for part (c). For a full statement and proof, see Feller (1968, p. 330). Theorem 4.2.2 is sufficient for our needs; a proof appears in Appendix B.

In the present case, $f = \sum_k p_k = 1$, $g_n = \delta_{n,0}$, and $v_n = b_n \mu^{-n}$. Thus, the Renewal Theorem implies that $\lim_{N \rightarrow \infty} b_N / \mu^N$ exists and equals $(\sum_{k=1}^{\infty} k p_k)^{-1}$. If this limit were strictly positive, then it would say that the expected "time between renewals" would be finite, and hence that a typical N -step bridge would consist of at least ϵN irreducible bridges for some $\epsilon > 0$. But this would imply that the average value of $|\omega(N)|$ over the set of N -step bridges ω beginning at 0 would be proportional to N rather than to N^ν . But scaling theory predicts that only an exponentially small fraction of N -step self-avoiding walks have $|\omega(N)| > \epsilon N$ [e.g. Fisher (1966)], which contradicts the hypothesis that $\lim_N b_N / \mu^N > 0$. Therefore it is believed that $\lim_N b_N / \mu^N = 0$, but there is no known proof of this.

So far, we have only counted bridges according to the x_1 coordinates of their sites. To obtain more detailed information, we will also want to look at the remaining $d-1$ coordinates. To this end, we introduce the following analogue of Definition 4.1.7 for irreducible bridges.

Definition 4.2.3 For each point y in $\mathbb{Z}^{d-1}$ and each positive integer L and N , let $\lambda_{N,L}(y)$ denote the number of N -step irreducible bridges ω with $\omega(0) = 0$ and $\omega(N) = (L, y)$, and let

$$\lambda_{N,L} = \sum_{y \in \mathbb{Z}^{d-1}} \lambda_{N,L}(y).$$

For each real $z > 0$, we define the generating functions

$$\Lambda_z(L, y) = \sum_{N=1}^{\infty} \lambda_{N,L}(y) z^N \quad \text{and} \quad \Lambda_z(L) = \sum_{N=1}^{\infty} \lambda_{N,L} z^N.$$

Using this definition, we have the following refinements of (4.2.2):

$$b_{N,L} = \sum_{k=1}^N \sum_{j=1}^L \lambda_{k,j} b_{N-k,L-j} + \delta_{N,0} \delta_{L,0} \quad (4.2.7)$$

$$b_{N,L}(y) = \sum_{k=1}^N \sum_{j=1}^L \sum_{v \in \mathbb{Z}^{d-1}} \lambda_{k,j}(v) b_{N-k,L-j}(y-v) + \delta_{N,0} \delta_{L,0} \delta_{y,0}. \quad (4.2.8)$$

These imply the following equations for the generating functions:

$$B_z(L) = \sum_{j=1}^L \Lambda_z(j) B_z(L-j) + \delta_{L,0} \quad (4.2.9)$$

$$B_z(L, y) = \sum_{j=1}^L \sum_{v \in \mathbb{Z}^{d-1}} \Lambda_z(j, v) B_z(L-j, y-v) + \delta_{L,0} \delta_{y,0}. \quad (4.2.10)$$

For the rest of this section, we shall only consider fixed $z < z_c$, so that $m(z) = M(z) = \overline{M}(z) > 0$ [by Theorem 4.1.14 and Proposition 4.1.1(b)]. First, let us turn (4.2.9) into a probabilistic renewal equation. We multiply both sides of the equation by $\exp(m(z)L)$, and set

$$p_L \equiv p_L(z) = \Lambda_z(L) e^{m(z)L} \quad (L \geq 1) \quad (4.2.11)$$

and

$$a_L \equiv a_L(z) = B_z(L) e^{m(z)L} \quad (L \geq 0). \quad (4.2.12)$$

Observe that $p_L \leq a_L \leq 1$ [by (4.1.12)]. Evidently the renewal equation (4.2.5) holds; to complete the probabilistic interpretation, we must show that $\sum_k p_k = 1$. Again, this can be accomplished by generating functions: take

$$P(s) = \sum_{k=1}^{\infty} p_k s^k \quad \text{and} \quad A(s) = \sum_{n=0}^{\infty} a_n s^n; \quad (4.2.13)$$

these are finite for $|s| < 1$. From (4.2.9) we then get

$$A(s) = \frac{1}{1 - P(s)} \quad \text{for } |s| < 1. \quad (4.2.14)$$

The sequence a_L is bounded away from zero [in fact, $a_L \geq (\chi(z))^{-2}$ by (4.2.12), (4.1.22), and (4.1.21)]. Therefore $A(s)$ diverges as s increases to 1, and so $P(1)$ must equal 1. That is,

$$\sum_{k=1}^{\infty} p_k = 1. \quad (4.2.15)$$

In addition, since we have just seen that a_L is bounded away from 0, the renewal theorem tells us that

$$\lim_{L \rightarrow \infty} a_L = \frac{1}{\sum_{k=1}^{\infty} k p_k} > 0. \quad (4.2.16)$$

Therefore

$$\sum_{k=1}^{\infty} k p_k < +\infty. \quad (4.2.17)$$

Using this kind of soft argument, there is not much more that can be said about the moments of the p_k sequence. However it turns out that

$$\limsup_{k \rightarrow \infty} p_k^{1/k} < 1. \quad (4.2.18)$$

This means that p_k has an exponential moment, or equivalently that the radius of convergence of $P(s)$ is strictly greater than 1. This can be expressed in terms of a "mass" for irreducible bridges:

Theorem 4.2.4 For $0 < z < z_c$, define

$$m_{\Lambda}(z) = \liminf_{L \rightarrow \infty} \frac{-\log \Lambda_z(L)}{L}. \quad (4.2.19)$$

Then

$$m_{\Lambda}(z) > m(z). \quad (4.2.20)$$

This theorem will be proven in the next section. It is clear that (4.2.20) is equivalent to (4.2.18).

We remark that it is not hard to see that (4.2.20) holds when z is sufficiently small. This is because an irreducible bridge of span $L \geq 2$ must have at least $3L$ steps, and when z is small the main contributions come from the shortest walks. Thus

$$\Lambda_z(L) = \sum_{N=3L}^{\infty} \lambda_{N,L} z^N \leq \sum_{N=3L}^{\infty} \mu^N z^N = \frac{(\mu z)^{3L}}{1 - \mu z}, \quad (4.2.21)$$

so

$$m_{\Lambda}(z) \geq -3 \log(\mu z). \quad (4.2.22)$$

But $m(z) \leq -\log z$ by Proposition 4.1.9(a), so $m_{\Lambda}(z) > m(z)$ whenever $z < \mu^{-3/2}$. Also note that these inequalities, together with Proposition 4.1.9(b) and the fact that

$$m_{\Lambda}(z) \leq -3 \log z \quad (4.2.23)$$

(which follows from the fact that $\lambda_{3L,L} \geq 1$ for $L \geq 2$), imply that $m_{\Lambda}(z) \sim 3m(z)$ as $z \searrow 0$. This is intuitively clear: for small z , the dominant contributions are from the shortest walks, and the shortest bridge of span L has length L while the shortest irreducible bridge has length $3L$. The $z \searrow 0$ limit is in fact where the greatest relative discrepancy between the two masses occurs, for we shall see in Corollary 4.4.4 that $m_{\Lambda}(z) \leq 3m(z)$ for every z in $(0, z_c)$.

For the rest of this section, we will concentrate on some of the consequences of Theorem 4.2.4. The first consequence is that the convergence in (4.2.16) is exponentially fast:

Theorem 4.2.5 For $0 < z < z_c$, there exists a strictly positive constant $\epsilon(z)$ such that

$$|B_z(L) e^{m(z)L} - C_z| \leq e^{-\epsilon(z)L}$$

for all $L \geq 1$, where $C_z = (\sum_{k=1}^{\infty} k \Lambda_z(k) e^{m(z)k})^{-1} > 0$.

Proof. Let s be complex, and define $P(s)$ and $A(s)$ as in (4.2.13). We know that $P(1) = 1$ by (4.2.15), and since the coefficients in the series defining $P(s)$ are all positive, we must have $|P(s)| < 1$ for all $s \neq 1$ such that $|s| \leq 1$. By Theorem 4.2.4, (4.2.18) holds, so $P(s)$ is analytic in some disc $|s| < R$ with R strictly greater than 1. Therefore there is an r between 1 and R such that $s = 1$ is the only zero of $P(s) - 1$ in the disc $|s| \leq r$. Next, observe that $P'(1) = 1/C_z$; in particular, $P'(1) \neq 0$, by (4.2.16). Denote the residue of the function $f(s)$ at a by $\text{Res}(f(s), a)$. By the residue theorem and the integral theorem for the coefficients of a Laurent series, we have

$$a_L = -\text{Res}\left(\frac{s^{-L-1}}{1 - P(s)}, 1\right) + \frac{1}{2\pi i} \oint_{|s|=r} \frac{s^{-L-1}}{1 - P(s)} ds.$$

The first term on the right is $1/P'(1)$ and the second is $O(r^{-L-1})$. This proves the theorem. $\square$

The probabilistic counterpart to the above theorem—that the probability of a renewal at time L converges exponentially fast to its limiting value when X_1 has an exponential moment—is well known (see Nummelin (1984), p.107). The next theorem is intrinsically probabilistic in nature: it gives a local central limit theorem for the endpoint of a bridge. In the terminology of mathematical physics, the $L^{-(d-1)/2}$ power law correction to the pure exponential decay in (4.2.25) below is known as *Ornstein-Zernike decay*. Such decay occurs in the subcritical two-point functions of a wide variety of statistical mechanical systems.

Theorem 4.2.6 For $0 < z < z_c$, there exists a strictly positive constant δ_z such that

$$\lim_{L \rightarrow \infty} \left| B_z(L, y) e^{m(z)L} L^{(d-1)/2} - C_z \frac{\exp(-|y|^2/\delta_z L)}{(\pi\delta_z)^{(d-1)/2}} \right| = 0 \quad (4.2.24)$$

uniformly in y in $\mathbf{Z}^{d-1}$. Consequently,

$$B_z(L, y) \sim C_z e^{-m(z)L} \frac{\exp(-|y|^2/\delta_z L)}{(\pi\delta_z L)^{(d-1)/2}} \quad (4.2.25)$$

as $L \rightarrow \infty$, uniformly in y in $\mathbf{Z}^{d-1}$ satisfying $|y| \leq KL^{1/2}$ (for every fixed $K > 0$). Equivalently we have (by Theorem 4.2.5)

$$\frac{B_z(L, y)}{B_z(L)} \sim \frac{\exp(-|y|^2/\delta_z L)}{(\pi\delta_z L)^{(d-1)/2}} \quad (4.2.26)$$

for the same range of y .

Intuitively, the relation (4.2.26) says that the endpoint of a bridge of span L has an asymptotically Gaussian probability distribution in the hyperplane $x_1 = L$.

We note here that Theorems 4.2.5 and 4.2.6 can be used to prove exactly analogous results for the full two-point function $G_z(0, x)$. This will be done in Section 4.4 (see Theorem 4.4.7).

We shall now provide a probabilistic context for interpreting and proving Theorem 4.2.6. Define the random vector (X, Y) such that X takes values in $\{1, 2, \dots\}$ and Y takes values in $\mathbf{Z}^{d-1}$, and

$$\Pr\{(X, Y) = (L, y)\} = \Lambda_z(L, y) e^{m(z)L} \text{ for } L \geq 1, y \in \mathbf{Z}^{d-1}. \quad (4.2.27)$$

Thus the marginal distribution of X is given by (4.2.11):

$$\Pr\{X = L\} = p_L(z), \text{ for } L = 1, 2, \dots$$

(In particular, (4.2.15) guarantees that (4.2.27) describes a genuine probability distribution.) Let $\{(X_n, Y_n) : n \geq 1\}$ be a sequence of independent copies of (X, Y) . Then we claim that

$$B_z(L, y) e^{m(z)L} = \Pr\{X_1 + \dots + X_k = L \text{ and } Y_1 + \dots + Y_k = y \text{ for some } k\} \quad (4.2.28)$$

for $L \geq 1$ and y in $\mathbf{Z}^{d-1}$. (This also holds for $L = 0$ if we interpret the event on the right as occurring for $k = 0$ when $(L, y) = (0, 0)$.) Equation (4.2.28) follows by iteration of (4.2.10), which gives

$$B_z(L, y) = \sum_{k=1}^L \left[\sum_{i=1}^k \prod_{i=1}^k \Lambda_z(n_i, v_i) \right]$$

where the inner sum is over all $n_1, \dots, n_k \geq 1$ and all $v_1, \dots, v_k$ in $\mathbf{Z}^{d-1}$ such that $n_1 + \dots + n_k = L$ and $v_1 + \dots + v_k = y$.

Now, if we let $S_n = \sum_{i=1}^n (X_i, Y_i)$, then $\{S_n\}$ is a d -dimensional random walk with positive, finite drift in the x_1 direction, and $B_z(L, y) e^{m(z)L}$ is the probability that this walk ever hits the point (L, y) . Thus Theorem 4.2.6 says that for large L , this probability factors into a $(d-1)$ -dimensional Gaussian density, with variance proportional to L , times the inverse of the mean of X [which is the limiting probability that the hyperplane $x_1 = L$ is ever hit, according to (4.2.16)]. This is just what one would expect probabilistically, if one knew that Y had finite variance. But in fact Theorem 4.2.4 implies that $|Y|$ has an exponential moment, as we now show.

Lemma 4.2.7 (a) $m_\Lambda(z)$ is a concave function of $\log z$ ($z > 0$). In particular, it is finite and continuous for $0 < z < z_c$.

(b) Fix $0 < z < z_c$. Then there exists an $s > 0$ for which $E(e^{s|Y|})$ is finite.

Proof. (a) By Lemma 4.1.2, $-L^{-1} \log \Lambda_z(L)$ is a concave function of $\log z$ for every $L \geq 1$. The desired concavity then follows from the fact that the lim inf of a concave sequence is concave. Inequalities (4.2.22) and (4.2.23) prove finiteness, and continuity follows immediately.

(b) By part (a), $m_\Lambda(z)$ is continuous at z . Therefore by (4.2.20) there exists an $s > 0$ such that $ze^s < z_c$ and $m_\Lambda(ze^s) > m(z)$. Then

$$\begin{aligned} E(e^{s|Y|}) &= \sum_{L, y} e^{s|y|} \Lambda_z(L, y) e^{m(z)L} \\ &\leq \sum_{L, y, N} e^{sN} \lambda_{N, L}(y) z^N e^{m(z)L} \\ &= \sum_L \Lambda_{ze^s}(L) e^{m(z)L} < \infty. \end{aligned}$$

□

We now know exactly why Theorem 4.2.6 should be true—all that is left is to complete the technical details. This was done in Chayes and Chayes (1986a) via an explicit asymptotic analysis of the generating functions. We shall follow a different route here, appealing directly to a theorem from the probability literature.

Proof of Theorem 4.2.6. The following result is a (very) special case of Theorem 3.2 of Stam (1971); we shall not reproduce its proof here. Let $\{(X_n, Y_n) : n \geq 1\}$ be an independent, identically distributed sequence of $\mathbf{Z}^d$ -valued random vectors (we follow our usual notation, so Y_n takes values

in $\mathbf{Z}^{d-1}$). We write the coordinates as follows:

$$(X_n, Y_n) = (X_n, Y_{n2}, Y_{n3}, \dots, Y_{nd}).$$

Assume the following: (i) that the expectation of X_1 , which we denote θ , is finite and strictly positive; (ii) that the $((d-1)/2)$ -th moment of X_1 is finite; (iii) that the covariance of Y_{1i} and Y_{1j} is 0 whenever $i \neq j$, and that for every $i = 2, \dots, d$ the variance of Y_{1i} is a strictly positive finite constant v ; and (iv) that the distribution of (X_n, Y_n) has no periodicities (for our purposes, it suffices to check that $\Pr\{(X_1, Y_1) = (L, y)\} > 0$ for every $L \geq 1$ and every y in $\mathbf{Z}^{d-1}$). Define

$$\varphi(L, y) = \theta^{-1} \left(\frac{\theta}{2\pi v} \right)^{(d-1)/2} \exp(-\theta|y|^2/2vL),$$

and let $U(L, y)$ denote the expected number of values of n such that

$$(X_1 + \dots + X_n, Y_1 + \dots + Y_n) = (L, y)$$

[i.e. the expected number of times that this d -dimensional random walk visits the point (L, y)]. Then Stam's theorem says that

$$\lim_{L \rightarrow \infty} |L^{(d-1)/2} U(L, y) - \varphi(L, y)| = 0,$$

uniformly in y in $\mathbf{Z}^{d-1}$. As a simple corollary, if we restrict $|y|$ to be no larger than $KL^{1/2}$ for some constant K , then $\varphi(L, y)$ remains bounded away from 0 and so

$$U(L, y) \sim L^{-(d-1)/2} \varphi(L, y)$$

as $L \rightarrow \infty$, uniformly for such y .

In our case, the distribution of the random vector is given by (4.2.27), for which assumption (iv) clearly holds. Since X_1 is always a positive integer, $\theta > 0$; in fact, $\theta = 1/C_z$, where C_z is the positive constant defined in Theorem 4.2.5. So assumption (i) holds. Assumption (ii) follows from the fact that X_1 has an exponential moment, by (4.2.18). For (iii): first, the covariances are 0 because the joint distribution of Y_{1i} and Y_{1j} is the same as the joint distribution of Y_{1i} and $-Y_{1j}$; secondly, symmetry implies that the variance of Y_{1i} does not depend on i ; and thirdly, v is finite by Lemma 4.2.7(b). Finally, since X_1 is strictly positive, a point which is visited once by the random walk is never revisited; thus (4.2.28) tells us that $U(L, y) = B_z(L, y)e^{m(z)L}$. Now Stam's theorem proves Theorem 4.2.6, with $\delta_z = 2v/\theta$. $\square$

As a final consequence of Theorem 4.2.4, we shall prove that the mass is a real analytic function of z in $(0, z_c)$. This will be accomplished by applying the analytic implicit function theorem to (4.2.15), which is an equation relating z and $m(z)$.

Theorem 4.2.8 *The mass $m(z)$ is a real analytic function of z in the interval $(0, z_c)$.*

Proof. Fix a real z_0 in $(0, z_c)$ and let $u_0 = e^{m(z_0)}$. Define the function f of two (complex) arguments u and z by

$$f(u, z) \equiv \sum_{L=1}^{\infty} \sum_{N=1}^{\infty} \lambda_{N,L} z^N u^L = \sum_{L=1}^{\infty} \Lambda_z(L) u^L. \quad (4.2.29)$$

We must first show that $f(u, z)$ is convergent (and hence holomorphic) in some neighbourhood of (u_0, z_0) . To this end, suppose that ϵ is a small positive number and suppose that (u, z) satisfies $|u - u_0| < \epsilon$ and $|z - z_0| < \epsilon$. Then

$$\sum_{L=1}^{\infty} \sum_{N=1}^{\infty} \lambda_{N,L} |z|^N |u|^L \leq \sum_{L=1}^{\infty} \Lambda_{z_0+\epsilon}(L) (e^{m(z_0)} + \epsilon)^L. \quad (4.2.30)$$

Applying the root test to the right hand sum and recalling the definition of m_Λ in (4.2.19), we see that the sum converges absolutely if

$$e^{-m_\Lambda(z_0+\epsilon)} (e^{m(z_0)} + \epsilon) < 1. \quad (4.2.31)$$

Since m_Λ is continuous at z_0 [by Lemma 4.2.7(a)] and since $m(z_0) - m_\Lambda(z_0) < 0$ (by Theorem 4.2.4), it is possible to choose ϵ small enough so that (4.2.31) holds. Therefore f is indeed holomorphic in a neighbourhood of (u_0, z_0) .

We see from Equations (4.2.29), (4.2.11), and (4.2.15) that

$$f(e^{m(z)}, z) = 1 \quad \text{for all } z \text{ in } (0, z_c); \quad (4.2.32)$$

in particular, $f(u_0, z_0) = 1$. Also, $\partial f / \partial u$ is nonzero at (u_0, z_0) , since it may be written as a series of positive terms. Therefore the analytic implicit function theorem [see for example Griffiths and Harris (1978), p. 19] tells us that there exists a function $w(z)$, holomorphic in a neighbourhood of z_0 , with the following property: in a neighbourhood of (u_0, z_0) , the equation $f(u, z) = 1$ holds if and only if $u = w(z)$. By (4.2.32) we see that $m(z) = \log w(z)$ for all real z in a neighbourhood of z_0 . This proves the theorem. $\square$

4.3 Separation of the masses

This section is devoted to a proof of Theorem 4.2.4, which states that the “mass” $m_\Lambda(z)$ of irreducible bridges is strictly greater than the mass $m(z)$ whenever $0 < z < z_c$.

Definition 4.3.1 Let $\omega = (\omega(0), \dots, \omega(N))$ be a bridge. A backtrack of ω is a subwalk

$$\omega[s; t] \equiv (\omega(s), \dots, \omega(t))$$

($0 \leq s < t \leq N$) satisfying:

- (i) $\omega_1(t) \leq \omega_1(i) < \omega_1(s)$ for all $i = s+1, \dots, t-1$;
- (ii) $\omega_1(j) \leq \omega_1(s)$ for all $j = 0, 1, \dots, s-1$; and
- (iii) $\omega_1(t) < \omega_1(j)$ for all $j = t+1, \dots, N$.

The span of the backtrack is defined to be $\omega_1(s) - \omega_1(t)$.

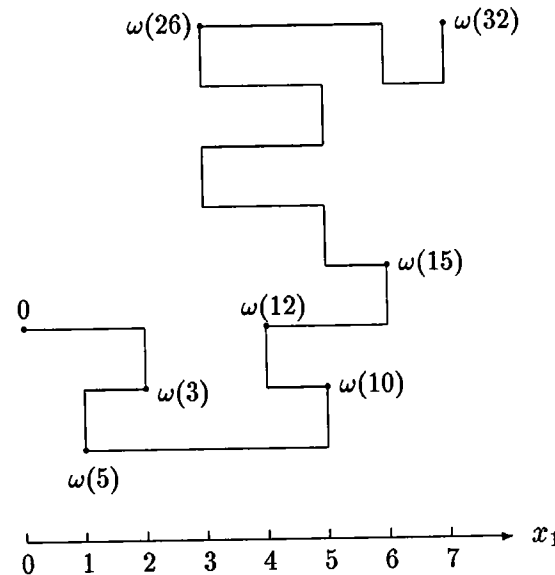


Figure 4.2: A bridge of length $N = 32$, with $\omega_1(0) = 0$ and $\omega_1(N) = 7$. There are three backtracks: $\omega[3; 5]$, $\omega[10; 12]$, and $\omega[15; 26]$.

Condition (i) says that a backtrack is itself a bridge, except that it goes right-to-left instead of left-to-right. Conditions (ii) and (iii) are maximality conditions: they guarantee that every subwalk $\omega[s; t]$ satisfying (i) is contained in a unique backtrack. In fact, the following is true:

4.3. SEPARATION OF THE MASSES

Lemma 4.3.2 Let ω be an N -step bridge. If $0 \leq s_0 < t_0 \leq N$ and $\omega_1(t_0) < \omega_1(s_0)$, then there is a unique backtrack $\omega[s, t]$ of ω that contains $\omega(s_0)$ and $\omega(t_0)$.

Proof. Given such s_0 and t_0 , define

$$\begin{aligned} A &= \min\{\omega_1(i) : t_0 \leq i \leq N\}, \\ B &= \max\{\omega_1(i) : 0 \leq i \leq s_0\}, \\ s &= \max\{i : 0 \leq i \leq s_0, \omega_1 = B\}, \\ t &= \max\{i : t_0 \leq i \leq N, \omega_1 = A\}. \end{aligned}$$

Then $\omega[s; t]$ is the unique backtrack containing $\omega(s_0)$ and $\omega(t_0)$. Details are left to the reader. $\square$

Corollary 4.3.3 Two backtracks of the same bridge are either equal or else they have no sites in common.

We now state two additional definitions related to the decomposition of a bridge into two smaller bridges, as can be done when (4.2.1) occurs for $i < N$.

Definition 4.3.4 Let ω be an N -step bridge. A backtrack $\omega[s; t]$ is said to cover the integer j if $\omega_1(t) \leq j < \omega_1(s)$. We say that the integer j [$\omega_1(0) < j < \omega_1(N)$] is a break point of ω if there exists an r in $\{1, \dots, N-1\}$ such that $\omega_1(i) \leq j$ for all $i = 0, \dots, r$ and $\omega_1(i) > j$ for all $i = r+1, \dots, N$.

For example, in Figure 4.2, the backtrack $\omega[15; 26]$ covers the integers 3, 4, and 5; and the integers 2 and 6 are break points (corresponding respectively to $r = 6$ and $r = 30$ in the definition). Notice that an integer can be covered by more than one backtrack (e.g. the integer 4 in Figure 4.2). Several remarks about Definition 4.3.4 are in order. Observe that j is a break point if and only if there exists an r such that $\omega[0; r]$ and $\omega[r; N]$ are both bridges, and $\omega_1(r) = j$. Also, j is a break point if and only if there is *only one* r such that $\omega_1(r) = j$ and $\omega_1(r+1) = j+1$. A bridge is irreducible if and only if it has no break points. Finally, for every j strictly between $\omega_1(0)$ and $\omega_1(N)$, either j is a break point or j is covered by a backtrack.

For the rest of this section, we consider a fixed value of z , with $0 < z < z_c$, so we shall usually suppress z in our notation (we shall write m for $m(z)$, $B(L)$ for $B_z(L)$, etc.). We will use the following notation: If $\mathcal{S}$ is a set of self-avoiding walks, then $GF(\mathcal{S})$ denotes the generating function of $\mathcal{S}$:

$$GF(\mathcal{S}) \equiv GF(\mathcal{S}, z) \equiv \sum_{\omega \in \mathcal{S}} z^{|\omega|}.$$

We begin with a basic lemma which says that break points are common in the (subcritical) ensemble of bridges. In particular, it says that a long interval on the x_1 -axis is unlikely to be free of break points. Later we will try to apply this bound to many long intervals simultaneously in order to prove that a complete absence of break points is exponentially rare, which is basically what Theorem 4.2.4 says.

Lemma 4.3.5 *For integers $c \geq 0$, $T \geq 1$ and $L \geq c + T$, let $B^*(L; c, T)$ denote the generating function of all bridges of span L with $\omega(0) = 0$ which have no break points among the integers $c + 1, \dots, c + T - 1$. Then there exists a decreasing function $\epsilon(T)$ (independent of L and c) such that $\lim_{T \rightarrow \infty} \epsilon(T) = 0$ and*

$$B^*(L; c, T) \leq e^{-mL} \epsilon(T) \text{ for all } L \text{ and } c \text{ (} 0 \leq c \leq L - T \text{)}. \quad (4.3.1)$$

[For example, the bridge of Figure 4.2 is in $B^*(7; 2, 4)$ since 3, 4, and 5 are not break points, but it is not in $B^*(7; 1, 4)$ or $B^*(7; 2, 5)$.]

Proof. Considering the last break point j of a bridge before $c + 1$ and its first break point k after $c + T - 1$, we see that

$$B^*(L; c, T) = \sum_{j=0}^c \sum_{k=c+T}^L B(j) \Lambda(k - j) B(L - k).$$

Using the notation of the previous section [recall (4.2.11) and (4.2.12)],

$$\begin{aligned} B^*(L; c, T) e^{mL} &= \sum_{j=0}^c \sum_{k=c+T}^L a_j p_{k-j} a_{L-k} \\ &\leq \sum_{i=T}^L (i - T + 1) p_i \end{aligned}$$

(where we have used $a_n \leq 1$ for all n). Therefore setting

$$\epsilon(T) = \sum_{i=T}^{\infty} i p_i$$

gives a function with the desired properties, by (4.2.17). $\square$

There is a “renewal theory” interpretation of the preceding lemma. The quantity $B^*(L; c, T) e^{mL}$ is the probability that there are no renewals between c and $c + T$ (and that there is a renewal at L). Since the mean time between renewals is finite, it is unlikely that a given long interval contains no renewals.

Our strategy now is the following. We want to bound the generating function of the set of irreducible bridges of span L (which start at the origin). We fix a large integer Q and for $L \gg Q$ we split the interval $[0, L]$ into blocks (subintervals) of size Q . Then we look at the backtracks that cover the endpoints of these blocks. For the subset of irreducible bridges in which many of these backtracks are small, the blocks approximately decouple into irreducible bridges, each of which has small probability (by Lemma 4.3.5), so many such blocks are exponentially rare. The remaining irreducible bridges have many large backtracks, so we use $B(k) \leq e^{-mk}$ to bound the contributions of these backtracks, again resulting in exponentially smaller quantities.

For the details, we proceed as follows. Let T and Δ be positive integers (to be specified in the proof of Theorem 4.2.4 below), and let $Q = 2\Delta + T$. For large L , let $k \equiv k(L)$ be the greatest integer less than or equal to $(L/Q) - 1$; we will split $[0, L]$ into $k + 1$ subintervals (observe that $k + 1$ is of order L). Given L , define the set

$$A = \{Q, 2Q, \dots, kQ\}.$$

With 0 and L , the elements of A are the endpoints of the blocks. Consider any nonempty subset $S = \{n_1, \dots, n_\tau\}$ of A , with $n_1 < \dots < n_\tau$. Put $n_0 = 0$ and $n_{\tau+1} = L$. Then

$$n_{i+1} - n_i \geq Q \text{ for all } i = 0, 1, \dots, \tau.$$

In the rest of this section, τ will always denote $|S|$. Let $\mathcal{I}_L$ denote the set of all irreducible bridges of span L . Let $\mathcal{I}_L(\leq \Delta; S)$ denote the set of all irreducible bridges of span L such that no point of S is covered by a backtrack having a span of more than Δ . For integers $\sigma_1, \dots, \sigma_\tau \geq 1$, let $\mathcal{J}_L(S, \sigma_1, \dots, \sigma_\tau)$ denote the set of all irreducible bridges of span L such that for each $i = 1, \dots, \tau$ there is a backtrack of span σ_i which covers n_i but does not cover n_j for any other $j \neq i$.

We will now state three lemmas. We will then show how they can be used to prove that $m_\Lambda > m$, and finally we will prove the lemmas. The first lemma says that every irreducible bridge either has lots of small backtracks (i.e. of span Δ or less) or it has enough large backtracks. The second lemma bounds the generating function of the first kind of irreducible bridge, and the third lemma does the same for the second kind. The notation of the preceding paragraphs is assumed in each lemma.

Lemma 4.3.6 *The set $\mathcal{I}_L$ is contained in the union of*

$$\bigcup_{S \subset A, |S| \geq k/2} \mathcal{I}_L(\leq \Delta; S) \quad (4.3.2)$$

and

$$\bigcup_{S \subset A, |S| \geq 1} \bigcup_{\substack{\sigma_1 > \Delta, \dots, \sigma_r > \Delta \\ \sigma_1 + \dots + \sigma_r > k\Delta/2}} \mathcal{I}_L(S, \sigma_1, \dots, \sigma_r). \quad (4.3.3)$$

Lemma 4.3.7 $GF(\mathcal{I}_L(\leq \Delta; S)) \leq e^{-mL} (\chi\epsilon(T))^{r+1}$.

Here ϵ is from Lemma 4.3.5 and χ is the susceptibility [defined in (1.3.4)].

Lemma 4.3.8 $GF(\mathcal{I}_L(S, \sigma_1, \dots, \sigma_r)) \leq e^{-mL} \chi^{2r} e^{-m(\sigma_1 + \dots + \sigma_r)}$.

Proof of Theorem 4.2.4. Fix T so that

$$2(\chi\epsilon(T))^{1/2} < \frac{1}{2}, \quad (4.3.4)$$

and fix Δ so that

$$\left(1 + \frac{\chi^2}{1 - e^{-m/2}}\right) e^{-m\Delta/4} < \frac{1}{2}. \quad (4.3.5)$$

Set $Q = 2\Delta + T$. Then, by the preceding three lemmas,

$$\begin{aligned} \Lambda_z(L) &\equiv GF(\mathcal{I}_L) \\ &\leq 2^k e^{-mL} [\chi\epsilon(T)]^{1+k/2} \\ &\quad + e^{-mL} \sum_{\tau=1}^k \binom{k}{\tau} \chi^{2\tau} \sum_{\substack{\sigma_1 > \Delta, \dots, \sigma_\tau > \Delta \\ \sigma_1 + \dots + \sigma_\tau > k\Delta/2}} e^{-m(\sigma_1 + \dots + \sigma_\tau)}. \end{aligned}$$

Now, for any positive D and r such that $r < m$,

$$\begin{aligned} &\sum_{\substack{\sigma_1 > \Delta, \dots, \sigma_r > \Delta \\ \sigma_1 + \dots + \sigma_r > D}} e^{-m(\sigma_1 + \dots + \sigma_r)} \\ &\leq \sum_{\sigma_1 > \Delta, \dots, \sigma_r > \Delta} e^{-m(\sigma_1 + \dots + \sigma_r)} e^{r(\sigma_1 + \dots + \sigma_r - D)} \\ &\leq \left(\frac{e^{(r-m)\Delta}}{1 - e^{-(m-r)}} \right)^r e^{-rD}. \end{aligned}$$

Putting $r = m/2$ and $D = k\Delta/2$, we obtain [using (4.3.4) and (4.3.5)]

$$\begin{aligned} \Lambda_z(L) &\leq e^{-mL} \left(\left[2(\chi\epsilon(T))^{1/2} \right]^k + \left[1 + \frac{\chi^2 e^{-m\Delta/2}}{1 - e^{-m/2}} \right]^k e^{-m\Delta k/4} \right) \\ &\leq e^{-mL} \left(\left(\frac{1}{2} \right)^k + \left(\frac{1}{2} \right)^k \right) \\ &\leq 2e^{-mL} 2^{2-(L/Q)}. \end{aligned}$$

Therefore

$$m_\Lambda = \liminf_{L \rightarrow \infty} \frac{-\log \Lambda_z(L)}{L} \geq m + \frac{\log 2}{Q} > m.$$

This proves the theorem. $\square$

Proof of Lemma 4.3.6. Suppose ω is in $\mathcal{I}_L$ but not in

$$\bigcup_{S \subset A, |S| \geq k/2} \mathcal{I}_L(\leq \Delta; S).$$

Then at least $k/2$ points of A are covered by backtracks of span greater than Δ . Some of these backtracks may cover more than one point of A . If a backtrack covers j points of A (where $j \geq 2$), then its span is at least $(j-1)Q$, which is greater than $j\Delta$ (recall $Q = 2\Delta + T$). Also, some points of A may be covered by several backtracks.

Choose an integer τ , as small as possible, having the following property: there exists a collection of τ backtracks of ω , each of span greater than Δ , such that at least $k/2$ of the points of A are covered by one or more of these τ backtracks. Then each backtrack covers some point of A which is not covered by any of the other $\tau-1$ backtracks (otherwise, this backtrack could be removed from the collection, contradicting the minimality of τ). This shows that for some $S = \{n_1, \dots, n_\tau\} \subset A$ and some $\sigma_1, \dots, \sigma_\tau > \Delta$ satisfying $\sigma_1 + \dots + \sigma_\tau > (k/2)\Delta$, ω is in $\mathcal{I}_L(S, \sigma_1, \dots, \sigma_\tau)$. This proves the lemma. $\square$

Proof of Lemma 4.3.7. Suppose that ω is in $\mathcal{I}_L(\leq \Delta; S)$, $S = \{n_1, \dots, n_\tau\}$. Let

$$\begin{aligned} r_0 &= 0, \\ q_{\tau+1} &= |\omega|, \\ q_i &= \min\{j : \omega_1(j+1) = n_i + 1\}, \quad i = 1, \dots, \tau, \\ r_i &= \max\{j : \omega_1(j) = n_i\}, \quad i = 1, \dots, \tau. \end{aligned}$$

(See Figure 4.3.) Notice that since ω has no break points, q_i is always strictly less than r_i . By our choice of ω , we know that

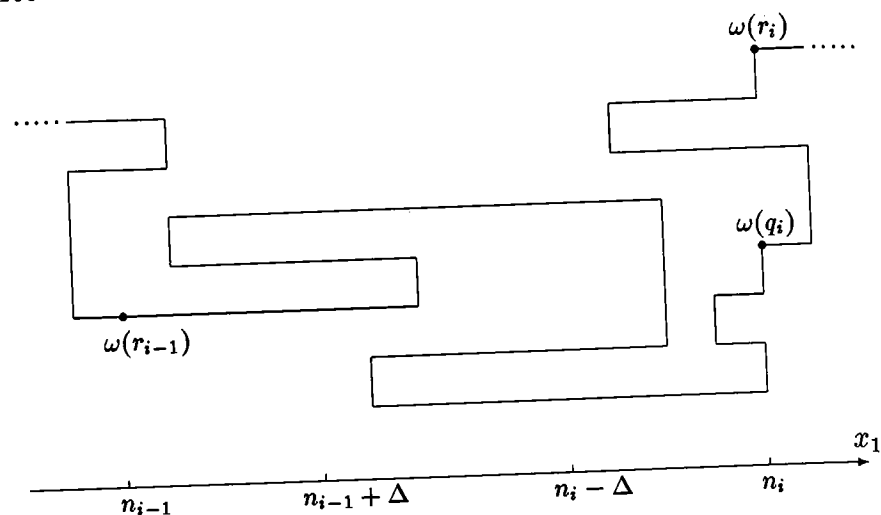
$$\omega_1(j) \geq n_i - \Delta \quad \text{for all } j \geq q_i$$

and

$$\omega_1(j) \leq n_i + \Delta \quad \text{for all } j \leq r_i.$$

In particular, $r_{i-1} < q_i$ ($i = 1, \dots, \tau+1$), since $n_{i+1} - n_i > 2\Delta$. Moreover, $\omega[r_{i-1}; q_i]$ is a bridge, and the rest of ω stays out of the strip

$$\{x \in \mathbb{R}^d : n_{i-1} + \Delta < x_1 < n_i - \Delta\}.$$

Figure 4.3: Proof of Lemma 4.3.7: part of a walk ω in $\mathcal{I}_L(\leq \Delta; S)$.

So, since ω has no break points, $\omega[r_{i-1}; q_i]$ has no break points between $n_{i-1} + \Delta$ and $n_i - \Delta$ (although it could have break points elsewhere). So, by cutting ω at each of $q_1, r_1, \dots, q_\tau, r_\tau$, and looking at the $2\tau + 1$ subwalks that are obtained, we deduce (with the help of Lemma 4.3.5) that

$$\begin{aligned} GF(\mathcal{I}_L(\leq \Delta; S)) &\leq \chi^\tau \prod_{i=1}^{\tau+1} B^*(n_i - n_{i-1}; \Delta, n_i - n_{i-1} - 2\Delta) \\ &\leq \chi^\tau \prod_{i=1}^{\tau+1} e^{-m(n_i - n_{i-1})} \epsilon(n_i - n_{i-1} - 2\Delta). \end{aligned}$$

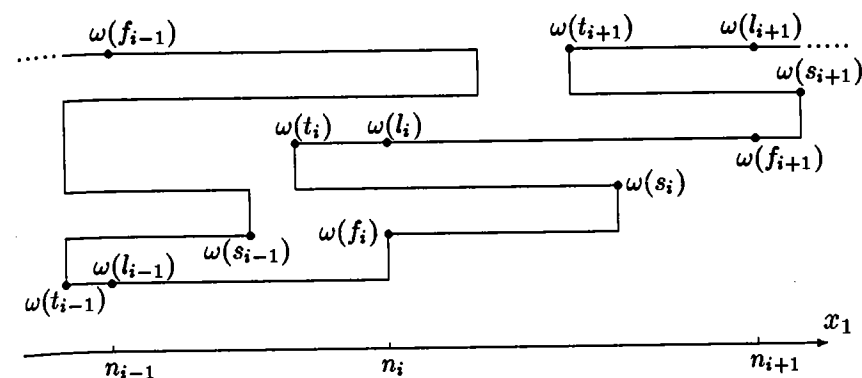
The lemma now follows because $\chi > 1$, $\sum_{i=1}^{\tau+1} (n_i - n_{i-1}) = L$, $n_i - n_{i-1} - 2\Delta \geq T$, and ϵ is a decreasing function. $\square$

Proof of Lemma 4.3.8. Suppose ω is in $\mathcal{J}_L(S, \sigma_1, \dots, \sigma_\tau)$. For every $i = 1, \dots, \tau$, there is at least one backtrack $\omega[s_i; t_i]$ of span σ_i such that

$$n_{i-1} < \omega_1(t_i) \leq n_i < \omega_1(s_i) \leq n_{i+1}.$$

(See Figure 4.4.) This implies that $t_i < s_{i+1}$ for every $i = 1, \dots, \tau - 1$. Define $l_0 = 0$ and, for $i = 1, \dots, \tau$,

$$\begin{aligned} f_i &= \min\{r > l_{i-1} : \omega_1(r+1) = n_i + 1\}, \\ l_i &= \max\{r : \omega_1(r) = n_i\}. \end{aligned}$$

Figure 4.4: Proof of Lemma 4.3.8: part of a walk ω in $\mathcal{J}_L(S, \sigma_1, \dots, \sigma_\tau)$.

Then, for every $i = 1, \dots, \tau$,

$$l_{i-1} < f_i < s_i < t_i \leq l_i,$$

and $\omega[l_{i-1}; f_i]$ is a bridge of span $n_i - n_{i-1}$. If we consider cutting ω at every f_i, s_i, t_i , and l_i , then we deduce that

$$\begin{aligned} GF(\mathcal{J}_L(S, \sigma_1, \dots, \sigma_\tau)) &\leq \left(\prod_{i=1}^{\tau+1} B(n_i - n_{i-1}) \right) \prod_{i=1}^{\tau} (\chi B(\sigma_i) \chi) \\ &\leq e^{-mL} \chi^{2\tau} e^{-m(\sigma_1 + \dots + \sigma_\tau)}, \end{aligned}$$

since $\sum_{i=1}^{\tau+1} (n_i - n_{i-1}) = L$. $\square$

4.4 Ornstein-Zernike decay of $G_z(0, x)$

The preceding section showed that for subcritical z , the spatial decay rate for irreducible bridges is strictly larger than for bridges as a whole, or in other words that the ratio $\Lambda_z(L)/B_z(L)$ decays exponentially in L for any fixed $z < z_c$. This may be viewed intuitively as saying that in the subcritical ensemble of bridges with distant endpoints, irreducible bridges are exponentially rare. (In contrast, bridges with distant endpoints are *not* exponentially rare among all self-avoiding walks with the same endpoints, by Theorem 4.1.14). The first part of this section proves some results in the opposite direction: we get a lower bound on the scarcity of irreducible bridges in the form $m_\Lambda(z) \leq 3m(z)$ whenever $0 < z < z_c$ (Corollary 4.4.4).

Since $m(z)$ tends to 0 as z approaches z_c , this tells us intuitively that irreducible bridges are not exponentially rare in the critical ensemble. We shall also prove a more natural interpretation of this intuitive statement, namely that the “connective constant” for irreducible bridges is the same as for walks (Corollary 4.4.5). Finally, we shall use these results to help us prove detailed large-distance asymptotics (the “Ornstein-Zernike” decay) of the subcritical two-point function $G_z(0, x)$ (Theorem 4.4.7).

Definition 4.4.1 For all nonnegative integers N , L , and r , let $\hat{B}_{N,L,r}$ be the set of bridges $(\omega(0), \dots, \omega(N))$ of span L such that $\omega(0) = 0$, $\omega_2(N) = r$, and $0 \leq \omega_2(i) \leq r$ for every $i = 0, \dots, N$. Also let $\hat{B}_{N,L} = \cup_{r \geq 0} \hat{B}_{N,L,r}$.

Such bridges are useful for the following reason, which we shall explain more carefully below: we produce an irreducible bridge whenever we perform an appropriate concatenation of three of them, all with the same span but with the middle one reflected in the x_1 direction so that it goes right-to-left. Thus to get lower bounds on the number of irreducible bridges, we first derive a lower bound on the number of bridges in $\hat{B}_{N,L}$.

Proposition 4.4.2 There exists an N_0 such that $b_{N,L} \leq e^{3N^{1/2}} |\hat{B}_{N,L}|$ for every $N \geq N_0$ and every $L \geq 1$.

Proof. The idea is very similar to the proof of Theorem 3.1.1, but now we will “unfold” the walk in the x_2 direction. Let ω be any N -step bridge of span L . Let $n_0(\omega)$ be the largest value of i such that $\omega_2(i) = \min_j \omega_2(j)$. For each $j > 0$, recursively define $A_j(\omega)$ and $n_j(\omega)$ so that

$$A_j = \max_{n_0 < i \leq N} (-1)^j (\omega_2(n_{j-1}) - \omega_2(i))$$

and n_j is the largest value of i for which this maximum is attained. The recursion is stopped at the smallest integer k such that $n_k = N$. Also, let $m_0(\omega) = n_0(\omega)$, and for each $j > 0$ recursively define $\bar{A}_j(\omega)$ and $m_j(\omega)$ so that

$$\bar{A}_j = \max_{0 \leq i < m_{j-1}} (-1)^j (\omega_2(m_{j-1}) - \omega_2(i))$$

and m_j is the smallest value of i for which this maximum is attained. The recursion is stopped at the smallest integer l such that $m_l = 0$. (See Figure 4.5.) Now, reflect the first m_{l-1} points of ω through the hyperplane $x_2 = \omega_2(m_{l-1})$, and continue reflecting through x_2 hyperplanes at $\omega(m_j)$, $j = l-2, \dots, 0$, and at $\omega(n_j)$, $j = 1, \dots, k-1$. The result of each reflection is still a bridge of span L , and the final result (after a translation in the x_2 direction and perhaps an overall reflection through $x_2 = 0$) will be in the set $\hat{B}_{N,L,A+\bar{A}}$, where $A = \sum A_j$ and $\bar{A} = \sum \bar{A}_j$. Since $A_1 > A_2 >$

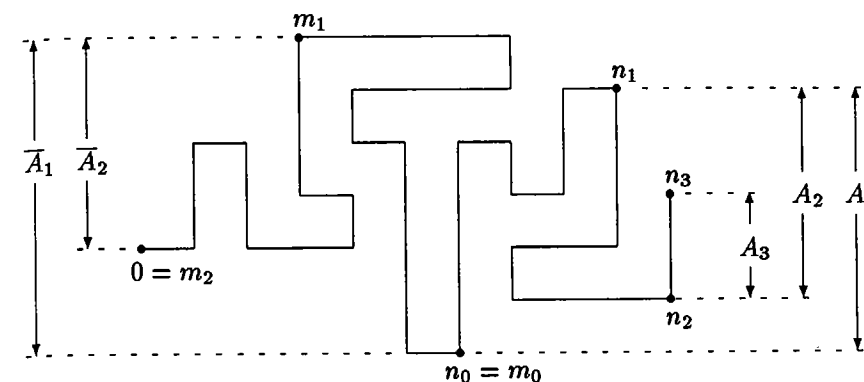


Figure 4.5: Proof of Proposition 4.4.2: The walk ω before “unfolding”.

$\dots > A_k$, there are $P_D(A)$ possible sequences of A_j 's that sum to A ; and since $\bar{A}_1 \geq \bar{A}_2 > \dots > \bar{A}_l$, there are $P_D(\bar{A} + 1)$ possible sequences of $\bar{A}_j$'s that sum to $\bar{A}$. Observe that these two sequences together with the final bridge in $\hat{B}_{N,L,A+\bar{A}}$ determine the original ω uniquely, except perhaps for a reflection through $x_2 = 0$. Therefore, since $A + \bar{A} \leq N$, we obtain

$$\begin{aligned} b_{N,L} &\leq 2 \sum_{r=0}^N \sum_{A=0}^r P_D(A) P_D(r-A+1) |\hat{B}_{N,L,r}| \\ &\leq 2 \sum_{r=0}^N (N+1) K^2 \exp \left[(3-\epsilon)(N+1)^{1/2} \right] |\hat{B}_{N,L,r}|, \end{aligned}$$

where the second inequality follows as in (3.1.7), having taken $B = 3$ in (3.1.6). The proposition follows. $\square$

We now establish several corollaries of this result. Recall that $M(z)$ is the mass for bridges, and that $M(z) = m(z)$ for all z in $(0, z_c)$.

Corollary 4.4.3 Let $\hat{B}_z(L) = \sum_N |\hat{B}_{N,L}| z^N$. Then

$$\lim_{L \rightarrow \infty} \frac{-\log \hat{B}_z(L)}{L} = M(z)$$

for every $z > 0$.

Proof. By Proposition 4.4.2 and the observation that bridges of span L must have at least L steps,

$$\hat{B}_z(L) \geq \sum_{N=L}^{\infty} b_{N,L} e^{-3N^{1/2}} z^N$$

for every $L \geq N_0$ and every $z > 0$. Now let $\epsilon > 0$. Choose $N_1 \geq N_0$ such that $e^{-3N^{1/2}} > (1-\epsilon)^N$ for every $N \geq N_1$. Then $\hat{B}_z(L) \geq B_{(1-\epsilon)z}(L)$ for every $L \geq N_1$, and hence

$$\limsup_{L \rightarrow \infty} \frac{-\log \hat{B}_z(L)}{L} \leq M((1-\epsilon)z). \quad (4.4.1)$$

Since ϵ is arbitrary and M is left-continuous [Lemma 4.1.11(d)], we can replace $M((1-\epsilon)z)$ by $M(z)$ on the right side of (4.4.1). The result then follows since the reverse inequality for the lim inf is a consequence of $\hat{B}_z(L) \leq B_z(L)$. $\square$

Corollary 4.4.4 $m_\Lambda(z) \leq 3m(z)$ for every z in $(0, z_c)$. Also, $m_\Lambda(z_c) = 0$.

Proof. Fix L , and let ω_A , ω_B and ω_C be three walks in $\cup_N \hat{B}_{N,L}$. Let $\bar{\omega}_B$ be the reflection of ω_B through the hyperplane $x_1 = 0$. Let ω be the walk obtained by starting at the origin, taking one step in the $+x_1$ direction, followed by ω_A (appropriately translated), followed by one step in the $+x_2$ direction, followed by $\bar{\omega}_B$, followed by one step in the $+x_2$ direction, followed by ω_C . Then ω is self-avoiding; in fact, it is an irreducible bridge of span $L+1$. This proves that $\Lambda_z(L+1) \geq [z\hat{B}_z(L)]^3$, and so Corollary 4.4.3 implies that $m_\Lambda(z) \leq 3M(z)$ for every $z > 0$. With Theorem 4.1.14, this proves the first assertion of the corollary. The second assertion follows from $M(z) \leq m_\Lambda(z) \leq 3M(z)$ and the fact that $M(z_c) = 0$ (Theorem 4.1.13). $\square$

Corollary 4.4.5 $\lim_{N \rightarrow \infty} (\lambda_N)^{1/N} = \mu$.

Proof. Since $\lambda_N \leq c_N$, it suffices to show that $\liminf_N (\lambda_N)^{1/N} \geq \mu$. In the proof of Corollary 4.4.4, suppose that ω_A , ω_B , and ω_C are all in $\hat{B}_{n-1,L}$; then the resulting irreducible bridge ω has $3n$ steps. Thus $\lambda_{3n,L+1}$ is bounded below by $|\hat{B}_{n-1,L}|^3$. Notice that we can get the same lower bound for $\lambda_{3n+1,L+1}$ and $\lambda_{3n+2,L+1}$, since we can add one or two steps in the $+x_2$ direction to the end of ω . Therefore

$$\lambda_{N,L+1} \geq |\hat{B}_{\lfloor N/3 \rfloor - 1, L}|^3$$

(where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x). Proposition 4.4.2 implies that

$$\lambda_{N,L+1} \geq (b_{\lfloor N/3 \rfloor - 1, L})^3 e^{-9(N/3)^{1/2}}.$$

We now sum this inequality over $L = 1, \dots, N$ and apply Hölder's inequality in the form

$$N^2 \sum_{i=1}^N a_i^3 \geq \left(\sum_{i=1}^N a_i \right)^3$$

(for $a_i \geq 0$). This gives

$$\lambda_N \geq \frac{(b_{\lfloor N/3 \rfloor - 1})^3}{N^2} e^{-9(N/3)^{1/2}}$$

for every N . Finally we take N -th roots of both sides and let N tend to infinity; the desired result is a consequence of Corollary 3.1.6. $\square$

For the remainder of this section, we need to extend the definition of *break point* in Definition 4.3.4 to an arbitrary N -step self-avoiding walk ω . Now we say that the integer j is a break point of ω if there exists an r in $\{1, \dots, N-1\}$ such that $\omega_1(i) \leq j$ for all $i = 0, \dots, r$ and $\omega_1(i) > j$ for all $i = r+1, \dots, N$. Observe that ω cannot have a break point unless $\omega_1(0) < \omega_1(N)$, and that each break point j must satisfy $\omega_1(0) \leq j < \omega_1(N)$. So the only real change in the definition is allowing $j = \omega_1(0)$. Notice that the two definitions coincide for bridges: if ω is a bridge then $\omega_1(0)$ cannot be a break point because we do not allow r to be 0.

Before proving the Ornstein-Zernike decay of $G_z(0, x)$, we require a lemma which establishes a decay rate for the point-to-plane generating functions of a new class of walks which is larger than the class of irreducible bridges. These walks end on a specified hyperplane $x_1 = L$, lie in the half-space $x_1 \leq L$, and have no break points.

Lemma 4.4.6 Suppose $0 < z < z_c$. For each integer $L \geq 0$, let $H_z(L)$ be the generating function of the class of self-avoiding walks $(\omega(0), \dots, \omega(N))$ such that: $\omega(0) = 0$; $\omega_1(N) = L$; $\omega_1(i) \leq L$ for every $i = 0, \dots, N$; and ω has no break points. Then

$$\liminf_{L \rightarrow \infty} \frac{-\log H_z(L)}{L} = m_\Lambda(z).$$

Proof. Since $H_z(L) \geq \Lambda_z(L)$ we have

$$\liminf_{L \rightarrow \infty} \frac{-\log H_z(L)}{L} \leq m_\Lambda(z),$$

so it remains to prove the reverse inequality.

Consider any ω in the class of walks that corresponds to $H_z(L)$. Let T be the largest i such that $\omega_1(i) = 0$. Let

$$J = \max\{\omega_1(i) : 0 \leq i \leq T\},$$

and let I be the first value of i for which this maximum is attained. Observe that the subwalk $(\omega(T), \dots, \omega(N))$ is a bridge of span L , and that if it has any break points, then they must all be in $\{1, 2, \dots, J-1\}$. We cut ω at I , T , and the last break point of $(\omega(T), \dots, \omega(N))$; this decomposition shows that

$$H_z(L) \leq \sum_{J=0}^L G_z(J)^2 \sum_{k=0}^{J-1} B_z(k) \Lambda_z(L-k),$$

where $G_z(J)$ was defined in (4.1.19).

Choose any number ρ such that $m(z) < \rho < m_\Lambda(z)$. By definition of $m_\Lambda(z)$, there exists an $R > 0$ such that $\Lambda_z(L) \leq Re^{-\rho j}$ for every $j \geq 1$. Using this, (4.1.21), (4.1.12) and Theorem 4.1.14, we obtain

$$\begin{aligned} H_z(L) &\leq \sum_{J=0}^L [\chi(z)^2 e^{-m(z)J}]^2 \sum_{k=0}^{J-1} e^{-m(z)k} R e^{-\rho(L-k)} \\ &\leq \chi(z)^4 \sum_{J=0}^L e^{-2m(z)J} J R e^{(\rho-m(z))J} e^{-\rho L} \\ &\leq \chi(z)^4 R L^2 e^{-\rho L} \end{aligned}$$

where we have used $\rho \leq 3m(z)$ (Corollary 4.4.4) in the last line. The result now follows because ρ can be made arbitrarily close to $m_\Lambda(z)$. $\square$

Finally we are ready to extend Theorems 4.2.5 and 4.2.6 to complete our picture of the long-distance asymptotics of the subcritical two-point function $G_z(0, x)$.

Theorem 4.4.7 Fix z with $0 < z < z_c$.

(a) There exist strictly positive, finite constants $\bar{c}(z)$ and $\bar{C}_z$ such that

$$|G_z(L) e^{m(z)L} - \bar{C}_z| \leq e^{-\bar{c}(z)L}$$

for all sufficiently large L .

(b) (Ornstein-Zernike decay) For δ_z as in Theorem 4.2.6,

$$\lim_{L \rightarrow \infty} \left| G_z(0, (L, y)) e^{m(z)L} L^{(d-1)/2} - \bar{C}_z \frac{\exp(-|y|^2/\delta_z L)}{(\pi \delta_z)^{(d-1)/2}} \right| = 0$$

uniformly in y in $\mathbb{Z}^{d-1}$.

The analogues of the asymptotic relations (4.2.25) and (4.2.26) also hold.

Proof. The basic idea is that in the subcritical ensemble, most walks with distant endpoints will have lots of break points; the difference between bridges and general walks in this respect is a matter of “boundary conditions” (at $x_1 = 0$ and $x_1 = L$). Our first job will be to show that walks with no break points are negligible. We shall then sum the remaining walks according to the locations of their first and last break points (which will typically be close to the endpoints). Then we apply Theorems 4.2.5 and 4.2.6 to the middle parts of the walks, which are bridges.

Throughout this proof we shall use m to denote $m(z)$.

(a) Let $G_z^*(L)$ be the generating function of all self-avoiding walks ω such that $\omega(0) = 0$, $\omega_1(N) = L$, and ω has no break points. We claim that the mass of $G_z^*(L)$ is strictly greater than m . The proof is similar in spirit to the proof of Lemma 4.4.6, so we shall be brief. For any walk ω contributing to $G_z^*(L)$, define T and J as in the proof of Lemma 4.4.6, and also define the analogous quantities $T' = \min\{i : \omega_1(i) = L\}$ and $J' = \min\{\omega_1(i) : T' \leq i \leq |\omega|\}$. The contribution of all walks which have either $J > L/3$ or $J' < 2L/3$ can be bounded by $2G_z([L/3])^2 G_z(L)$, and hence has mass at least $5m/3$ by Corollary 4.1.17. The contribution of the remaining walks can be bounded by

$$\sum_{J=0}^{L/3} G_z(J)^2 \sum_{J'=2L/3}^L G_z(L-J')^2 \sum_{k=0}^J \sum_{k'=0}^{J'} B_z(k) \Lambda_z(k'-k) B_z(L-k'),$$

which also has mass strictly greater than m . This proves the claim. (We remark that with more care, one can show that the mass of $G_z^*(L)$ is $m_\Lambda(z)$.)

By considering the leftmost break point i and the rightmost break point $L-1-j$ of a given walk, we have

$$G_z(L) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1-i} H_z(i) B_z(L-1-i-j) z H_z(j) + G_z^*(L) \quad (4.4.2)$$

(the factor of z is due to the single step of the walk from $x_1 = L-1-j$ to $x_1 = L-j$). Define $\bar{C}_z = z e^m C_z (\sum_{n \geq 0} H_z(n) e^{nm})^2$, where C_z is defined in Theorem 4.2.5, and the sum in parentheses converges by Lemma 4.4.6 and Theorem 4.2.4. From the above identity we have

$$G_z(L) e^{Lm} - \bar{C}_z =$$

$$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1-i} z H_z(i) [B_z(L-1-i-j) e^{(L-1-i-j)m} - C_z] H_z(j) e^{(i+j+1)m} \\ - C_z \sum_{i+j \geq L} z H_z(i) H_z(j) e^{(i+j+1)m} + G_z^*(L) e^{Lm}.$$

Part (a) of the theorem now follows from Theorem 4.2.5 and Lemma 4.4.6.

(b) The proof is similar to part (a), but we need to pay more attention to technicalities. For every (L, y) in $\mathbb{Z}^d$, let $H_z(L, y)$ and $G_z^*(0, (L, y))$ denote the analogues of $H_z(L)$ and $G_z^*(L)$ when attention is restricted to walks that end at (L, y) . Then

$$G_z(0, (L, y)) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1-i} \sum_{u, v \in \mathbb{Z}^{d-1}} H_z(i, u) B_z(L-1-i-j, y-u-v) z H_z(j, v) \\ + G_z^*(0, (L, y)). \quad (4.4.3)$$

Since the last term is less than $G_z^*(L)$ it decays faster than e^{-Lm} . Therefore, after substituting $k = i+j+1$ in (4.4.3), we are left with the task of showing that

$$\sup_{y \in \mathbb{Z}^{d-1}} \left| \sum_{k=1}^L \sum_{i=0}^{k-1} \sum_{u, v} z H_z(i, u) B_z(L-k, y-u-v) H_z(k-1-i, v) e^{Lm} L^{(d-1)/2} \right. \\ \left. - z e^m \left(\sum_{n=0}^{\infty} H_z(n) e^{nm} \right)^2 \varphi(L, y) \right| \quad (4.4.4)$$

converges to 0 as L tends to infinity, where as in the proof of Theorem 4.2.6

$$\varphi(L, y) = C_z (\pi \delta_z)^{-(d-1)/2} \exp(-|y|^2 / \delta_z L).$$

As we shall see, this converges to 0 because the mass of the H_z terms is strictly greater than m and because

$$B_z(L, y) e^{Lm} L^{(d-1)/2} - \varphi(L, y) \rightarrow 0$$

uniformly in y as $L \rightarrow \infty$. To use these facts, we add and subtract several terms in (4.4.4) and use the triangle inequality, as well as the bound $B_z(L-k, y-u-v) e^{(L-k)m} \leq 1$ [which follows from (4.1.12) and Theorem 4.1.14]. As a result, (4.4.4) is bounded by

$$\sum_{k=1}^{L/2} \sum_{i=0}^{k-1} \sum_{u, v} z H_z(i, u) H_z(k-1-i, v) e^{km} (q_{L,k} + r_{L-k} + s_{L,k,u+v})$$

$$+ \sum_{k=L/2+1}^L \sum_{i=0}^{k-1} \sum_{u, v} z H_z(i, u) H_z(k-1-i, v) e^{km} L^{(d-1)/2} \\ + \sum_{k=L/2+1}^{\infty} \sum_{i=0}^{k-1} z H_z(i) H_z(k-i-1) e^{km} \varphi(L, 0),$$

where we define

$$q_{L,k} = \sup_{x \in \mathbb{Z}^{d-1}} |L^{(d-1)/2} - (L-k)^{(d-1)/2}| B(L-k, x) e^{(L-k)m},$$

$$r_n = \sup_{x \in \mathbb{Z}^{d-1}} |n^{(d-1)/2} B_z(n, x) e^{nm} - \varphi(n, x)|,$$

and

$$s_{L,k,w} = \sup_{y \in \mathbb{Z}^{d-1}} |\varphi(L-k, y-w) - \varphi(L, y)|.$$

By Lemma 4.4.6, there exist positive constants A and ϵ (depending only on z) such that

$$\sum_{i=0}^{k-1} \sum_{u, v} H_z(i, u) H_z(k-1-i, v) e^{km} = \sum_{i=0}^{k-1} H_z(i) H_z(k-1-i) e^{km} \\ \leq A e^{-\epsilon k} \quad (4.4.5)$$

for every $k \geq 1$. Thus, if we can prove that $q_{L,k}$, r_{L-k} , and $s_{L,k,w}$ (i) are bounded uniformly for all $L \geq 2k$ and all w , and (ii) converge to 0 as $L \rightarrow \infty$ for every fixed k and w , then the dominated convergence theorem will imply that (4.4.4) converges to 0, which is what we want.

To begin with, $\lim_{n \rightarrow \infty} r_n = 0$ by Theorem 4.2.6, so (i) and (ii) hold for r_{L-k} . Next we consider $q_{L,k}$. The uniform convergence in Theorem 4.2.6 implies that there is a finite constant Ψ (depending only on z and d) such that

$$(L-k)^{(d-1)/2} B_z(L-k, x) e^{(L-k)m} \leq \Psi$$

for every L, k , and x . This in turn implies that

$$q_{L,k} \leq \left| \left(\frac{L}{L-k} \right)^{(d-1)/2} - 1 \right| \Psi. \quad (4.4.6)$$

This proves (ii) for $q_{L,k}$. Since $0 \leq k \leq L/2$, the right side of (4.4.6) is bounded by $(2^{(d-1)/2} - 1)\Psi$, which proves (i).

Finally, φ is uniformly bounded, hence so is $s_{L,k,w}$. This proves (i) for this sequence. To check (ii), we need to consider bounds on

$$|\exp(-|y|^2/\delta L) - \exp(-|y-w|^2/\delta(L-k))|$$

that are uniform in y , where δ is a positive constant. First, the mean value theorem gives

$$\begin{aligned} |\exp(-|y|^2/\delta L) - \exp(-|y-w|^2/\delta L)| &\leq \sup_{t \in \mathbb{R}} \left| \frac{d}{dt} e^{-t^2/\delta} \right| \left| \frac{|y|}{\sqrt{L}} - \frac{|y-w|}{\sqrt{L}} \right| \\ &\leq \sup_{t \in \mathbb{R}} \left| \frac{2t}{\delta} e^{-t^2/\delta} \right| \frac{|w|}{\sqrt{L}} \\ &\leq \text{const.} \frac{|w|}{\sqrt{L}}. \end{aligned}$$

Secondly, for arbitrary $\zeta \geq 0$ (so in particular for $\zeta = |y-w|^2/\delta$):

$$\begin{aligned} |\exp(-\zeta/L) - \exp(-\zeta/(L-k))| &= \exp(-\zeta/L) |1 - \exp(-\zeta k/L(L-k))| \\ &\leq \exp(-\zeta/L) \frac{\zeta k}{L(L-k)} \\ &\leq \left(\sup_{t \geq 0} t e^{-t} \right) \frac{k}{L-k} \end{aligned}$$

where we have used $1 - e^{-a} \leq a$ for $a \geq 0$. The above two inequalities show that

$$s_{L,k,w} \leq \text{const.} \left(\frac{|w|}{\sqrt{L}} + \frac{k}{L-k} \right),$$

which proves (ii). □

4.5 Notes

Section 4.1. Chayes and Chayes (1986a) perform a systematic study of the mass, deriving many of the results that are discussed in this section. In particular, our presentation of the material from Proposition 4.1.8 through Corollary 4.1.17 is largely based upon their work. They also prove that $M(z) = -\infty$ for all $z > z_c$.

Theorem 4.1.18 is modelled upon an analogous result for percolation in Alexander, Chayes and Chayes (1990). In place of the FKG inequality used in that paper, we require Lemma 4.1.4, which we believe has not been used before for this purpose. The proof of Lemma 4.1.4 is closely related

to the proof of Lemma 1.5.2 (see in particular the first inequality depicted in Figure 1.2).

Section 4.2. Ornstein-Zernike decay of two-point correlation functions occurs in many lattice spin systems and Euclidean quantum field theories [see Chayes and Chayes (1986b) for references], as well as in percolation [Campanino, Chayes and Chayes (1991)]. The original reference [Ornstein and Zernike (1914)] was in the context of fluid mechanics.

The results of this section are due to Chayes and Chayes (1986a, 1986b). They give the Ornstein-Zernike decay with an explicit error term as a power of L^{-1} that is uniform over a certain region of $y \in \mathbb{Z}^{d-1}$ [see also the proof of Theorem 4.4 of Campanino, Chayes and Chayes (1991) for a correction of a misstatement in this respect in Chayes and Chayes (1986b)].

Section 4.3. Theorem 4.2.4 is due to Chayes and Chayes (1986b). Their proof splits the interval $[0, L]$ into blocks and uses a result similar to our Lemma 4.3.5, but then it works with various rescaled block generating functions and an “Ornstein-Zernike inequality”. The proof that we give in this section is from Madras (1991a).

Section 4.4. Theorem 4.4.7 is due to Chayes and Chayes (1986b), which they state in a different form that includes an explicit error term, as they did for bridges. This paper also partially anticipates Corollary 4.4.4 and Lemma 4.4.6. The other results of this section are new.

Chapter 5

The lace expansion

5.1 Inclusion-exclusion

So far the lace expansion is the only method which has led to rigorous results proving existence of critical exponents for the self-avoiding walk. All results obtained so far are for dimensions greater than the conjectured upper critical dimension four.

The lace expansion was first introduced by Brydges and Spencer (1985), who used it to study *weakly* self-avoiding walk above four dimensions. Weakly self-avoiding walk will be defined precisely in the next section; roughly speaking it is a model of a random walk where walks which intersect themselves do have a nonzero weight, but this weight is smaller than for a walk which does not intersect itself. The size of the probability penalty imposed for a self-intersection (i.e. the weakness of the interaction) is the small parameter which provided convergence for the lace expansion.

Now consider a model of self-avoiding walk on a lattice with large coordination number Z , or in other words a walk for which there are a large number of steps available at each site. The probability that the concatenation of a randomly chosen single step to a given walk produces an immediate reversal is Z^{-1} , which is small. For example, consider the usual nearest-neighbour model in high dimensions, with coordination number $2d$. As we saw in Section 1.2, in high dimensions the main effect of the self-avoidance constraint is in some sense to rule out immediate reversals, and now it appears that immediate reversals are uncommon even in the absence of the constraint. This suggests that we can regard the interaction as being weak in high dimensions. Alternatively, we may consider a walk on a “low”-dimensional lattice with large coordination number. Our basic example of

this type is a walk on the hypercubic lattice $\mathbf{Z}^d$, in which we allow all steps whose largest component has absolute value less than or equal to L , for some large parameter L . Then again the effect of ruling out immediate reversals will be small, but we must also worry about long-range effects. It will turn out that these are also small for $d > 4$, so that in this situation the self-avoidance is again weak. We shall see in Chapter 6 that in these two situations it is possible to obtain convergence of the lace expansion. In the former the small parameter responsible for convergence of the expansion is $(2d)^{-1}$, while in the latter it is L^{-1} . Remarkably, for the nearest-neighbour model in five dimensions the small parameter $1/10$ is sufficiently small to prove convergence.

The lace expansion was first derived by an expansion and resummation procedure reminiscent of the cluster expansions of statistical mechanics and constructive quantum field theory. Viewed differently, however, the lace expansion can be seen as resulting from repeated application of the inclusion-exclusion relation. In this section the derivation of the lace expansion via the inclusion-exclusion relation will be discussed, and in the next section the resummation approach will be described. The inclusion-exclusion approach is geometric and is useful for providing intuition as to the origin of the expansion, while the resummation procedure has the advantage of generating terms in the expansion in a systematic and algebraic manner. In the inclusion-exclusion approach, the name lace expansion may seem somewhat inappropriate, but the laces will appear in the next section. Both derivations lead to precisely the same expansion.

At the heart of the lace expansion method is a convolution equation for the two-point function which is a multi-dimensional analogue of the renewal equations of Section 4.2. Indeed the expansion amounts to identifying an "irreducible" two-point function, which will be denoted $\Pi_z(0, x)$, such that the two-point function is essentially given by its convolution with the irreducible two-point function. Fourier transform techniques will play a key role in the analysis of the convolution equation.

We now turn to the derivation of the expansion for the case of a walk which may take more general steps than just to a nearest neighbour. We fix a finite set $\Omega \not\ni 0$ in $\mathbf{Z}^d$ which is symmetric with respect to reflections in the coordinate hyperplanes and rotations about coordinate axes by $\pi/2$. The cardinality of Ω will also be denoted by Ω . We consider the ordinary random walk taking steps in Ω . More precisely, we consider $\omega = (\omega(0), \omega(1), \dots, \omega(N))$, where each $\omega(i)$ is an element of $\mathbf{Z}^d$ and $\omega(i+1) - \omega(i) \in \Omega$. (If Ω is the set of nearest neighbours of the origin then this is just the simple random walk.) It is shown in (A.6) that for $z < \Omega^{-1}$ the

Fourier transform of the two-point function for this walk is given by

$$\hat{C}_z(k) = \frac{1}{1 - z\Omega\hat{D}(k)},$$

where

$$\hat{D}(k) = \frac{1}{\Omega} \sum_{x \in \Omega} e^{ik \cdot x}.$$

To simplify the notation we will not use a label Ω to keep track of the fact that we are not necessarily dealing with the nearest-neighbour walk; all walks in this section will take steps in Ω . Self-avoiding walks taking steps in Ω will satisfy the same subadditivity inequality as the nearest-neighbour self-avoiding walk, and hence will have a critical point $z_c = z_c(\Omega)$. We define $\hat{\Pi}_z(k)$, for $z < z_c$, implicitly by the equation

$$\hat{G}_z(k) = \frac{1}{1 - z\Omega\hat{D}(k) - \hat{\Pi}_z(k)}. \quad (5.1.1)$$

Then $\hat{\Pi}_z$ can be thought of as a measure of the difference between self-avoiding and simple random walk. The lace expansion is an expansion for $\hat{\Pi}_z$.

We denote by $\mathcal{C}_N(x, y)$ the set of all N -step self-avoiding walks from x to y (taking steps in Ω), and denote its cardinality by $c_N(x, y)$. The first step in deriving the expansion is to extract the term in $G_z(0, x)$ corresponding to $N = 0$:

$$G_z(0, x) = \delta_{0,x} + \sum_{N=1}^{\infty} c_N(0, x) z^N. \quad (5.1.2)$$

We shall now argue that for $N \geq 1$,

$$c_N(0, x) = \sum_{y \in \Omega} \left[c_1(0, y) c_{N-1}(y, x) - \sum_{\omega^{(1)} \in \mathcal{C}_{N-1}(y, x)} I[0 \in \omega^{(1)}] \right]. \quad (5.1.3)$$

Diagrammatically the right side of (5.1.3) can be represented by

$$\sum_{y \in \Omega} \left[0 \text{ --- } \underset{y}{\text{---}} \text{---} x \quad - \quad 0 \text{ --- } \bigcirc \underset{y}{\text{---}} x \right].$$

In the first term on the right side the bold line is unconstrained, apart from the fact that it should be self-avoiding. The thin line in the first term

represents a single step. Equation (5.1.3) is just the inclusion-exclusion relation: the first term on the right side counts all walks from 0 to x which are self-avoiding *after* the first step, and the second subtracts the contribution due to those which are not self-avoiding from the beginning, i.e., walks that return to the origin. Since $c_1(0, y) = 1$ for $y \in \Omega$, substitution of (5.1.3) into (5.1.2) gives

$$G_z(0, x) = \delta_{0,x} + z \sum_{y \in \Omega} G_z(y, x) - \sum_{y \in \Omega} \sum_{N=0}^{\infty} z^{N+1} \sum_{\omega^{(1)} \in \mathcal{C}_N(y, x)} I[0 \in \omega^{(1)}]. \quad (5.1.4)$$

The inclusion-exclusion relation can now be applied to the last term on the right side of (5.1.4), as follows. Let S be the first (and only) time that $\omega^{(1)}(S) = 0$. Then

$$\begin{aligned} \sum_{\omega^{(1)} \in \mathcal{C}_N(y, x)} I[0 \in \omega^{(1)}] &= \sum_{S=1}^N \sum_{\substack{\omega^{(2)} \in \mathcal{C}_S(y, 0) \\ \omega^{(3)} \in \mathcal{C}_{N-S}(0, x)}} I[\omega^{(2)} \cap \omega^{(3)} = \{0\}] \\ &= \sum_{S=1}^N \left[c_S(y, 0) c_{N-S}(0, x) - \sum_{\substack{\omega^{(2)} \in \mathcal{C}_S(y, 0) \\ \omega^{(3)} \in \mathcal{C}_{N-S}(0, x)}} I[\omega^{(2)} \cap \omega^{(3)} \neq \{0\}] \right]. \end{aligned}$$

We can interpret $c_S(y, 0)$ as the number of $(S+1)$ -step walks which step from the origin directly to y , then return to the origin in S steps, which have distinct vertices apart from the fact that they return to their starting point. Let $\mathcal{U}_S$ denote the set of all S -step self-avoiding loops at the origin (S -step walks which begin and end at the origin but which otherwise have distinct vertices), and let u_S be the cardinality of $\mathcal{U}_S$. Then

$$\begin{aligned} \sum_{y \in \Omega} \sum_{N=0}^{\infty} z^{N+1} \sum_{\omega^{(1)} \in \mathcal{C}_N(y, x)} I[0 \in \omega^{(1)}] \\ = \sum_{S=2}^{\infty} z^S u_S \cdot G_z(0, x) - \sum_{\substack{S=2 \\ N=0}}^{\infty} \sum_{\substack{\omega^{(2)} \in \mathcal{U}_S \\ \omega^{(3)} \in \mathcal{C}_N(0, x)}} z^{S+N} I[\omega^{(2)} \cap \omega^{(3)} \neq \{0\}]. \end{aligned}$$

Continuing in this fashion, in the last term on the right side of the above equation let $T_1 \geq 1$ be the first time along $\omega^{(3)}$ that $\omega^{(3)}(T_1) \in \omega^{(2)}$, and let $v = \omega^{(3)}(T_1)$. Then the inclusion-exclusion relation can be applied

again to remove the avoidance between the portions of $\omega^{(3)}$ before and after T_1 , and correct for this removal by the subtraction of a term involving a further intersection. For $z < z_c$ repetition of this procedure leads to the convolution equation

$$G_z(0, x) = \delta_{0,x} + z \sum_{y \in \Omega} G_z(y, x) + \sum_v \Pi_z(0, v) G_z(v, x), \quad (5.1.5)$$

where the "irreducible" two-point function $\Pi_z(0, x)$ is given by

$$\Pi_z(0, v) = \sum_{N=1}^{\infty} (-1)^N \Pi_z^{(N)}(0, v), \quad (5.1.6)$$

with the terms on the right side defined as follows. The $N = 1$ term is given by

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

The $N = 2$ term is

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

where $I(\omega_1, \omega_2, \omega_3)$ is equal to 1 if the ω_i are pairwise mutually avoiding apart from their common endpoints, and otherwise equals 0. Diagrammatically this can be represented by

$$\Pi_z^{(2)}(0, v) = 0 \text{ } \bigcirc \text{ } v \text{ } ,$$

where each line represents a sum over self-avoiding walks between the endpoints of the line, weighted by z^T , with mutual avoidance between the three pairs of lines in the diagram. Similarly

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

where now there is mutual avoidance between some but not all pairs of lines in the diagram; we defer any discussion of the details of this mutual avoidance until later in the chapter. The unlabelled vertex is summed over

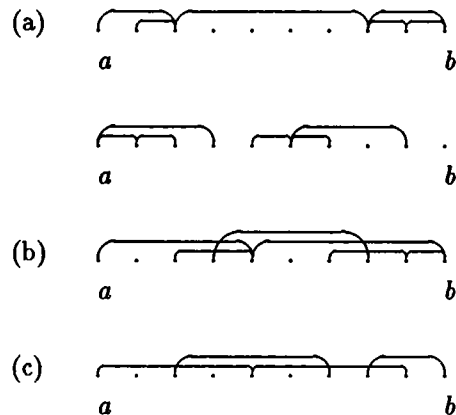


Figure 5.1: Graphs in which an edge st is represented by an arc joining s and t . (a) Examples of graphs which are not connected. (b) An example of a connected graph. (c) An example of a lace.

Definition 5.2.1 Given an interval $I = [a, b]$ of positive integers, we refer to a pair $\{s, t\}$ ($s < t$) of elements of I as an edge. To abbreviate the notation, we usually write st for $\{s, t\}$. The length of an edge st is $t - s$. A set of edges is called a graph. A graph Γ is said to be connected if both a and b are endpoints of edges in Γ , and if in addition, for any $c \in (a, b)$, there are $s, t \in [a, b]$ such that $s < c < t$ and $st \in \Gamma$. The set of all graphs on $[a, b]$ consisting of edges of length τ or less is denoted $\mathcal{B}_\tau[a, b]$, and the subset consisting of all connected graphs is denoted $\mathcal{G}_\tau[a, b]$. A lace is a minimally connected graph, i.e., a connected graph for which the removal of any edge would result in a disconnected graph. The set of laces on $[a, b]$ consisting of edges of length τ or less is denoted by $\mathcal{L}_\tau[a, b]$, and the set of laces on $[a, b]$ which consist of exactly N edges is denoted $\mathcal{L}_{\tau, N}[a, b]$.

A convenient graphical representation of graphs and laces is illustrated in Figure 5.1.

Given a connected graph Γ , the following prescription associates to Γ a unique lace $\mathcal{L}_\Gamma$: The lace $\mathcal{L}_\Gamma$ consists of edges $s_1 t_1, s_2 t_2, \dots$ where

$$s_1 = a, \quad t_1 = \max\{t : at \in \Gamma\}$$

$$t_{i+1} = \max\{t : st \in \Gamma, s < t_i\}$$

$$s_i = \min\{s : st_i \in \Gamma\}.$$

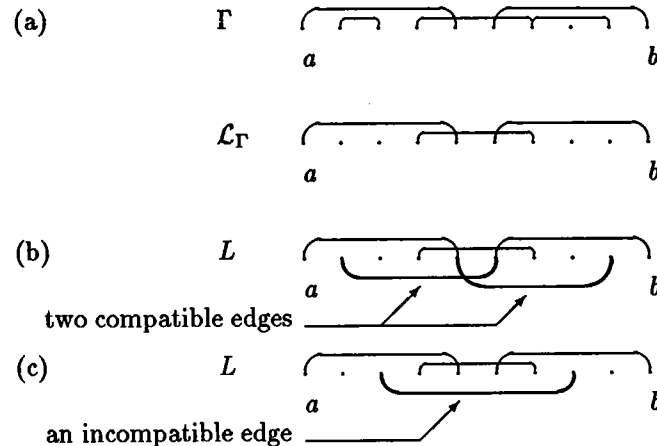


Figure 5.2: (a) An example of a connected graph Γ and its associated lace $L = \mathcal{L}_\Gamma$. (b) Examples of edges compatible with the lace L . (c) An example of an edge which is not compatible with the lace L .

Given a lace L , the set of all edges $st \notin L$ (of length τ or less) such that $\mathcal{L}_{L \cup \{st\}} = L$ is denoted $\mathcal{C}_\tau(L)$. Edges in $\mathcal{C}_\tau(L)$ are said to be *compatible* with L . Figure 5.2 illustrates these definitions.

For $a < b$ we define

$$K_\tau[a, b] = \prod_{\substack{s, t \in [a, b] \\ 0 < t - s \leq \tau}} (1 + \mathcal{U}_{st}). \quad (5.2.6)$$

We set $K_\tau[a, a] = 1$, and if $a > b$ then we set $K_\tau[a, b] = 0$. By expanding the product in (5.2.6) we obtain

$$K_\tau[a, b] = \sum_{\Gamma \in \mathcal{B}_\tau[a, b]} \prod_{st \in \Gamma} \mathcal{U}_{st}. \quad (5.2.7)$$

For $a < b$ we define an analogous quantity, in which the sum over graphs is restricted to connected graphs:

$$J_\tau[a, b] = \sum_{\Gamma \in \mathcal{G}_\tau[a, b]} \prod_{st \in \Gamma} \mathcal{U}_{st}. \quad (5.2.8)$$

We set $J_\tau[a, a] = 1$, and if $a > b$ then we set $J_\tau[a, b] = 0$. Partially

resumming the right side of (5.2.8), we obtain

$$\begin{aligned} J_\tau[a, b] &= \sum_{L \in \mathcal{L}_\tau[a, b]} \sum_{\Gamma: \mathcal{L}_\Gamma = L} \prod_{st \in L} \mathcal{U}_{st} \prod_{s't' \in \Gamma \setminus L} \mathcal{U}_{s't'} \\ &= \sum_{L \in \mathcal{L}_\tau[a, b]} \prod_{st \in L} \mathcal{U}_{st} \prod_{s't' \in \mathcal{C}_\tau(L)} (1 + \mathcal{U}_{s't'}). \end{aligned} \quad (5.2.9)$$

For $a < b$ we define $J_{\tau, N}[a, b]$ to be the contribution to (5.2.9) from laces consisting of exactly N bonds:

$$J_{\tau, N}[a, b] = \sum_{L \in \mathcal{L}_{\tau, N}[a, b]} \prod_{st \in L} \mathcal{U}_{st} \prod_{s't' \in \mathcal{C}_\tau(L)} (1 + \mathcal{U}_{s't'}). \quad (5.2.10)$$

Each term in the above sum is either 0 or $(-1)^N$. By (5.2.9) and (5.2.10),

$$J_\tau[a, b] = \sum_{N=1}^{\infty} J_{\tau, N}[a, b]. \quad (5.2.11)$$

The sum over N in (5.2.11) is a finite sum, since the sum in (5.2.10) is empty for $N > b - a$ and hence $J_{\tau, N}[a, b] = 0$ if $N > b - a$. By definition $J_\tau[a, a+1] = 0$, since the only lace on $[a, a+1]$ consists of the single edge $\{a, a+1\}$, and $\mathcal{U}_{a, a+1}(\omega) = 0$ for all ω , because a walk cannot be at the same place at consecutive times.

Lemma 5.2.2 *For any $a < b$,*

$$K_\tau[a, b] = K_\tau[a+1, b] + \sum_{j=a+2}^b J_\tau[a, j] K_\tau[j, b]. \quad (5.2.12)$$

Proof. The contribution to the sum on the right side of (5.2.7) due to all graphs Γ for which a is not in an edge is exactly $K_\tau[a+1, b]$. To resum the contribution due to the remaining graphs we proceed as follows. If Γ does contain an edge ending at a , let $j(\Gamma)$ be the largest value of j such that the set of edges in Γ with at least one end in the interval $[a, j]$ forms a connected graph on $[a, j]$. We lose nothing by taking $j \geq a+2$, since as argued above the statement of the lemma, $\mathcal{U}_{a, a+1} = 0$. Then resummation over graphs on $[j, b]$ gives

$$K_\tau[a, b] = K_\tau[a+1, b] + \sum_{j=a+2}^b \sum_{\Gamma \in \mathcal{G}_\tau[a, j]} \prod_{st \in \Gamma} \mathcal{U}_{st} K_\tau[j, b], \quad (5.2.13)$$

which with (5.2.8) proves the lemma. $\square$

Now we define

$$\Pi_z^{(N)}(0, x; \tau) = (-1)^N \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} J_{\tau, N}[0, |\omega|] \quad (5.2.14)$$

and

$$\Pi_z(0, x; \tau) = \sum_{N=1}^{\infty} (-1)^N \Pi_z^{(N)}(0, x; \tau) = \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} J_\tau[0, |\omega|], \quad (5.2.15)$$

for any z for which the right side converges. The factor $(-1)^N$ on the right side of (5.2.14) ensures that

$$\Pi_z^{(N)}(0, x; \tau) \geq 0 \quad \text{for nonnegative } z. \quad (5.2.16)$$

Since $J_{\tau, N}[0, 1] = 0$, the sum on the right side of (5.2.14) or (5.2.15) could equally well be over walks ω of length greater than or equal to 2. Recall the notation

$$\hat{D}(k) = \frac{1}{\Omega} \sum_{x \in \Omega} e^{ik \cdot x}$$

Theorem 5.2.3 *For any value of z for which $\sum_{|\omega| \geq 2} z^{|\omega|} J_\tau[0, |\omega|]$ and $\sum_{|\omega| \geq 0} z^{|\omega|} K_\tau[0, |\omega|]$ converge absolutely,*

$$G_z(0, x; \tau) = \delta_{0, x} + z \sum_{u \in \Omega} G_z(u, x; \tau) + \sum_v \Pi_z(0, v; \tau) G_z(v, x; \tau) \quad (5.2.17)$$

and

$$\hat{G}_z(k; \tau) = \frac{1}{1 - z\Omega\hat{D}(k) - \hat{\Pi}_z(k; \tau)}. \quad (5.2.18)$$

Proof. It suffices to obtain (5.2.17), since then (5.2.18) follows immediately upon taking the Fourier transform of (5.2.17) and using the fact that the Fourier transform of a convolution is the product of Fourier transforms. Existence of the Fourier transforms of $G_z(0, \cdot; \tau)$ and $\Pi_z(0, \cdot; \tau)$ is guaranteed by the hypotheses of the theorem.

To prove (5.2.17), we first extract the contribution to (5.2.2) due to the zero step walk:

$$G_z(0, x; \tau) = \delta_{0, x} + \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} K_\tau[0, |\omega|].$$

$$G_z(0, x; \tau) = \delta_{0,x} + \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} K_\tau[1, |\omega|] \\ + \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} \sum_{j=2}^{|\omega|} J_\tau[0, j] K_\tau[j, |\omega|]. \quad (5.2.19)$$

In the second term on the right side, the factor $K_\tau[1, |\omega|]$ is nonzero only if the walk is self-avoiding after the first step. Explicitly summing over the endpoint of the first step in ω , the second term is equal to

$$z \sum_{y \in \Omega} G_z(y, x; \tau).$$

The third term on the right side of (5.2.19) is equal to

$$\sum_{N=1}^{\infty} z^N \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| = N}} \sum_{j=2}^N J_\tau[0, j] K_\tau[j, N].$$

Interchanging the order of summation gives the following expression for this quantity:

$$\sum_{j=2}^{\infty} z^j \sum_{N=j}^{\infty} z^{N-j} \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| = N}} J_\tau[0, j] K_\tau[j, N].$$

In the sum over ω , there is no interaction between the initial j -step portion and final $(N - j)$ -step portion of the walk. Factorizing the walk into these two pieces gives the desired result.

In the last step we interchanged two infinite sums. This is justified by the hypothesis that these two sums converge absolutely. $\square$

Next, we prove an identity which will be used to study the finite-dimensional distributions of the self-avoiding walk. It expresses the difference between a self-avoiding walk and two independent or decoupled self-avoiding walks, in the spirit of the inclusion-exclusion relation. First we need a definition.

Definition 5.2.4 Any graph $B \in \mathcal{B}_\tau[0, b]$ breaks up into connected components in a natural way. Given an integer m in the open interval $(0, b)$, we let $C_m(B) = \{m\}$ if B does not contain a bond st with $s < m < t$. If B

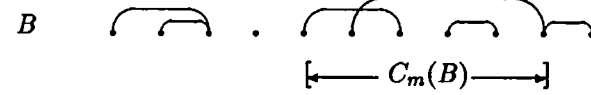


Figure 5.3: An example of a graph B and interval $C_m(B)$.

does contain a bond st with $s < m < t$, then there is a connected component Γ of B with bonds having endpoints less than and greater than m . In this case we let $C_m(B) = [i, j]$, where i is the smallest endpoint of all bonds in Γ , and j is the largest.

An example illustrating the definition is depicted in Figure 5.3.

Lemma 5.2.5 For any integers $0 \leq m \leq b$,

$$K_\tau[0, b] = \sum_{I \ni m} K_\tau[0, I_1] J_\tau[I_1, I_2] K_\tau[I_2, b],$$

where the sum over I is a sum over intervals $[I_1, I_2]$ of integers with either $0 \leq I_1 < m < I_2 \leq b$ or $I_1 = I_2 = m$.

Proof. By definition, $C_m(B)$ is an interval of the type being summed over in the statement of the lemma. Therefore a partial resummation of (5.2.7) gives

$$K_\tau[0, b] = \sum_{I \ni m} \sum_{B: C_m(B)=I} \prod_{st \in B} U_{st}.$$

Factoring the sum over B into three independent sums over graphs on $[0, I_1]$, connected graphs on $I = [I_1, I_2]$, and graphs on $[I_2, b]$, gives

$$K_\tau[0, b] = \sum_{I \ni m} K_\tau[0, I_1] \sum_{\Gamma \in \mathcal{G}_\tau(I)} \prod_{st \in \Gamma} U_{st} K_\tau[I_2, b].$$

The lemma then follows from (5.2.8). $\square$

Finally we prove a lemma which will be used in Section 6.7 to prove existence of the infinite self-avoiding walk above four dimensions. To state the lemma we first introduce some notation. Since we will not need a finite memory in Section 6.7, we consider only the fully self-avoiding walk, and drop subscripts τ . Given $n \geq m \geq 0$ and an n -step self-avoiding walk ω , let ω_m denote the first m steps of ω . For $m \geq 1$, we write $\mathbf{k} = (k^{(1)}, \dots, k^{(m)})$,

$$\hat{\Pi}_z(k; 2) = \frac{z^2 \Omega}{z^2 - 1} [1 - z \hat{D}(k)].$$

Proof. We use (5.2.15), (5.2.11) and (5.2.10), with $\tau = 2$, to evaluate $\hat{\Pi}_z(k; 2)$. When the memory is equal to two, all contributing laces have all edges of length exactly equal to two. There is a unique N -edge lace L_N contributing to $J_{2,N}[0, |\omega|]$ if $|\omega| = N + 1$, illustrated in Figure 5.4. If $|\omega| \neq N + 1$, then $J_{2,N}[0, |\omega|] = 0$. There is therefore exactly one term in the sum (5.2.10) defining $J_{2,N}[0, |\omega|]$ if $|\omega| = N + 1$, and there are no terms otherwise.

Thus we can write the Fourier transform of (5.2.14) as

$$\begin{aligned} \hat{\Pi}_z^{(N)}(k; 2) &= \sum_x \Pi_z^{(N)}(0, x; 2) e^{ik \cdot x} \\ &= (-1)^N z^{N+1} \sum_{|\omega|=N+1} J_{2,N}[0, N+1] e^{ik \cdot \omega(N+1)} \\ &= (-1)^N z^{N+1} \sum_{|\omega|=N+1} \prod_{st \in L_N} U_{st}(\omega) e^{ik \cdot \omega(N+1)}, \end{aligned} \quad (5.3.1)$$

where L_N is the unique memory-two N -edge lace. The product over compatible edges in (5.2.10) is equal to 1 here, since all bonds compatible with L_N have length one. The product $\prod_{st \in L_N} U_{st}(\omega)$ is equal to $(-1)^N$ if it is nonzero, and it will be nonzero if and only if ω has the topology indicated in Figure 5.4. These walks end at the origin if N is odd, and end at a site in Ω if N is even. In either case they simply step back and forth repeatedly between the origin and a particular site in Ω .

As an example of a calculation using the lace expansion, we now solve the self-avoiding walk with memory equal to two by finding an explicit formula for $\hat{G}_z(k; 2)$. This allows for an explicit calculation of the mean-square displacement. Although the calculation of the mean-square displacement for the memory-two walk is much simpler than the corresponding calculation for the fully self-avoiding walk, the memory-two calculation does illustrate some of the basic features which will occur in Chapter 6.

We begin with a formula for $\hat{\Pi}_z(k; 2)$. The derivation of the formula requires $|z| < 1$, but the resulting expression has an analytic continuation to a meromorphic function, which in turn provides a meromorphic extension for $\hat{G}_z(k; 2)$.

Theorem 5.3.1 *For the self-avoiding walk with memory equal to two,*

$$\hat{\Pi}_z(k; 2) = \frac{z^2 \Omega}{z^2 - 1} [1 - z \hat{D}(k)].$$

Proof. We use (5.2.15), (5.2.11) and (5.2.10), with $\tau = 2$, to evaluate $\hat{\Pi}_z(k; 2)$. When the memory is equal to two, all contributing laces have all edges of length exactly equal to two. There is a unique N -edge lace L_N contributing to $J_{2,N}[0, |\omega|]$ if $|\omega| = N + 1$, illustrated in Figure 5.4. If $|\omega| \neq N + 1$, then $J_{2,N}[0, |\omega|] = 0$. There is therefore exactly one term in the sum (5.2.10) defining $J_{2,N}[0, |\omega|]$ if $|\omega| = N + 1$, and there are no terms otherwise.

Thus we can write the Fourier transform of (5.2.14) as

$$\begin{aligned} \hat{\Pi}_z^{(N)}(k; 2) &= \sum_x \Pi_z^{(N)}(0, x; 2) e^{ik \cdot x} \\ &= (-1)^N z^{N+1} \sum_{|\omega|=N+1} J_{2,N}[0, N+1] e^{ik \cdot \omega(N+1)} \\ &= (-1)^N z^{N+1} \sum_{|\omega|=N+1} \prod_{st \in L_N} U_{st}(\omega) e^{ik \cdot \omega(N+1)}, \end{aligned} \quad (5.3.1)$$

where L_N is the unique memory-two N -edge lace. The product over compatible edges in (5.2.10) is equal to 1 here, since all bonds compatible with L_N have length one. The product $\prod_{st \in L_N} U_{st}(\omega)$ is equal to $(-1)^N$ if it is nonzero, and it will be nonzero if and only if ω has the topology indicated in Figure 5.4. These walks end at the origin if N is odd, and end at a site in Ω if N is even. In either case they simply step back and forth repeatedly between the origin and a particular site in Ω .

As an example of a calculation using the lace expansion, we now solve the self-avoiding walk with memory equal to two by finding an explicit formula for $\hat{G}_z(k; 2)$. This allows for an explicit calculation of the mean-square displacement. Although the calculation of the mean-square displacement for the memory-two walk is much simpler than the corresponding calculation for the fully self-avoiding walk, the memory-two calculation does illustrate some of the basic features which will occur in Chapter 6.

We begin with a formula for $\hat{\Pi}_z(k; 2)$. The derivation of the formula requires $|z| < 1$, but the resulting expression has an analytic continuation to a meromorphic function, which in turn provides a meromorphic extension for $\hat{G}_z(k; 2)$.

Theorem 5.3.1 *For the self-avoiding walk with memory equal to two,*

$$\hat{\Pi}_z(k; 2) = \frac{z^2 \Omega}{z^2 - 1} [1 - z \hat{D}(k)].$$

Proof. We use (5.2.15), (5.2.11) and (5.2.10), with $\tau = 2$, to evaluate $\hat{\Pi}_z(k; 2)$. When the memory is equal to two, all contributing laces have all edges of length exactly equal to two. There is a unique N -edge lace L_N contributing to $J_{2,N}[0, |\omega|]$ if $|\omega| = N + 1$, illustrated in Figure 5.4. If $|\omega| \neq N + 1$, then $J_{2,N}[0, |\omega|] = 0$. There is therefore exactly one term in the sum (5.2.10) defining $J_{2,N}[0, |\omega|]$ if $|\omega| = N + 1$, and there are no terms otherwise.

Thus we can write the Fourier transform of (5.2.14) as

$$\begin{aligned} \hat{\Pi}_z^{(N)}(k; 2) &= \sum_x \Pi_z^{(N)}(0, x; 2) e^{ik \cdot x} \\ &= (-1)^N z^{N+1} \sum_{|\omega|=N+1} J_{2,N}[0, N+1] e^{ik \cdot \omega(N+1)} \\ &= (-1)^N z^{N+1} \sum_{|\omega|=N+1} \prod_{st \in L_N} U_{st}(\omega) e^{ik \cdot \omega(N+1)}, \end{aligned} \quad (5.3.1)$$

where L_N is the unique memory-two N -edge lace. The product over compatible edges in (5.2.10) is equal to 1 here, since all bonds compatible with L_N have length one. The product $\prod_{st \in L_N} U_{st}(\omega)$ is equal to $(-1)^N$ if it is nonzero, and it will be nonzero if and only if ω has the topology indicated in Figure 5.4. These walks end at the origin if N is odd, and end at a site in Ω if N is even. In either case they simply step back and forth repeatedly between the origin and a particular site in Ω .

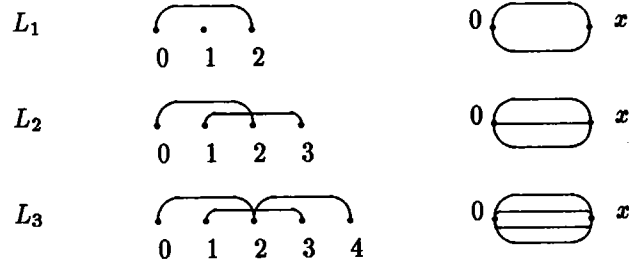


Figure 5.4: The laces for the memory-two walk, together with the corresponding walk topologies for $\prod_{s \in L_N} \mathcal{U}_{st} \neq 0$. Each line in the diagrams on the right represents a single step of the walk, and x is a site in Ω .

Summing (5.3.1) times $(-1)^N$ over N , with the odd and even values of N summed separately, gives

$$\begin{aligned} \hat{\Pi}_z(k; 2) &= -\sum_{n=1}^{\infty} z^{2n} \Omega + \sum_{n=1}^{\infty} z^{2n+1} \sum_{x \in \Omega} e^{ik \cdot x} \\ &= \frac{z^2}{z^2 - 1} \Omega [1 - z \hat{D}(k)]. \end{aligned} \quad (5.3.2)$$

□

Corollary 5.3.2 For all $z \in \mathbb{C}$,

$$\hat{G}_z(k; 2) = \frac{1 - z^2}{1 + (\Omega - 1)z^2 - z\Omega \hat{D}(k)}. \quad (5.3.3)$$

Proof. Combining Theorem 5.3.1 with Theorem 5.2.3, we obtain (5.3.3) for all $|z| < \mu_2^{-1} = (\Omega - 1)^{-1}$. But since the right side defines a function meromorphic in the plane, we have a meromorphic extension of $\hat{G}_z(k; 2)$. □

The denominator of (5.3.3) has zeroes

$$z_{\pm}(k) = \frac{\Omega \hat{D}(k) \pm [\Omega^2 \hat{D}(k)^2 - 4\Omega + 4]^{1/2}}{2(\Omega - 1)}. \quad (5.3.4)$$

At $k = 0$ these reduce to

$$z_-(0) = \frac{1}{\Omega - 1}, \quad z_+(0) = 1. \quad (5.3.5)$$

5.3. EXAMPLE: THE MEMORY-TWO WALK

For $k = 0$ the singularity at $z_+(0) = 1$ in (5.3.3) is removable, and hence the susceptibility $\hat{G}_z(0; 2)$ has a unique singularity at the simple pole $z = (\Omega - 1)^{-1}$. For k near but not equal to 0, $\hat{G}_z(k; 2)$ has two simple poles. The location of the closest singularity of $\hat{G}_z(0; 2)$ to the origin could have been anticipated from the fact that $z_c(0; 2)$, given in (5.2.5), is equal to μ_2^{-1} .

Let $c_{T,2}(0, x)$ denote the number of T -step memory-two walks ending at x , and denote its Fourier transform by $\hat{c}_{T,2}(k)$. Then the mean-square displacement of the memory-two walk is given by

$$\langle |\omega(T)|^2 \rangle_{T=2} = -\frac{\nabla_k^2 \hat{c}_{T,2}(0)}{\Omega(\Omega - 1)^{T-1}}, \quad (5.3.6)$$

where ∇_k denotes the gradient in k -space.

Let C be a small circle centred at the origin of the complex plane. Since

$$\hat{G}_z(k; 2) = \sum_{T=0}^{\infty} \hat{c}_{T,2}(k) z^T, \quad (5.3.7)$$

we have

$$-\nabla_k^2 \hat{c}_{T,2}(0) = \frac{1}{2\pi i} \oint_C \frac{-\nabla_k^2 \hat{G}_z(0; 2)}{z^{T+1}} dz. \quad (5.3.8)$$

By (5.3.3),

$$-\nabla_k^2 \hat{G}_z(0; 2) = \frac{(1 + z)z \sum_{y \in \Omega} |y|^2}{[1 - (\Omega - 1)z]^2(1 - z)}. \quad (5.3.9)$$

The integral on the right side of (5.3.8) can be evaluated exactly using the residue theorem. We deform the contour of integration past the singularities of the integrand at $(\Omega - 1)^{-1}$ and 1 to a large circle of radius R , and then let R go to infinity. The integral over the large circle vanishes in the limit, leaving the contributions from the residues. Denoting the residue of $f(z)$ at z_0 by $\text{Res}(f(z), z_0)$, we then have

$$\begin{aligned} \frac{1}{2\pi i} \oint_C \frac{-\nabla_k^2 \hat{G}_z(0; 2)}{z^{T+1}} dz \\ = -\text{Res} \left(\frac{-\nabla_k^2 \hat{G}_z(0; 2)}{z^{T+1}}, \frac{1}{\Omega - 1} \right) - \text{Res} \left(\frac{-\nabla_k^2 \hat{G}_z(0; 2)}{z^{T+1}}, 1 \right). \end{aligned} \quad (5.3.10)$$

To abbreviate the notation, let $b^2 = \Omega^{-1} \sum_{y \in \Omega} |y|^2$ and $\delta = (\Omega - 1)^{-1}$. Computing the residues and using (5.3.6) and (5.3.8) then gives

$$\langle |\omega(T)|^2 \rangle_{T=2} = b^2 \left[\left(\frac{1 + \delta}{1 - \delta} \right)^T - \frac{2\delta(1 - \delta^T)}{(1 - \delta)^2} \right]$$

$$\sim b^2 \left(\frac{1+\delta}{1-\delta} \right) T. \quad (5.3.11)$$

Thus the mean-square displacement is as expected asymptotically linear in the number of steps.

5.4 Bounds on the lace expansion

Equation (5.2.15) can be used to provide a diagrammatic representation for $\Pi_z(0, x; \tau)$ like that obtained using the inclusion-exclusion relation in Section 5.1. In this section we describe this diagrammatic representation, and then use it to obtain upper bounds on $\hat{\Pi}_z(k; \tau)$ and some of its derivatives. The bounds we obtain will be in terms of norms of the two-point function. Although we drop the memory τ from the notation, the results of this section are valid for any memory $2 \leq \tau \leq \infty$. The activity z is a complex parameter.

By (5.2.15), (5.2.14) and (5.2.10),

$$\begin{aligned} \Pi_z(0, x) &= \sum_{N=1}^{\infty} (-1)^N \Pi_z^{(N)}(0, x) \\ &= \sum_{N=1}^{\infty} \sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} \sum_{L \in \mathcal{L}_N[0, |\omega|]} \prod_{st \in L} \mathcal{U}_{st} \prod_{s't' \in \mathcal{C}(L)} (1 + \mathcal{U}_{s't'}). \end{aligned} \quad (5.4.1)$$

In this section we do not concern ourselves with the convergence or divergence of the above series, or of other related series which will occur. Rather, we treat these series as formal power series in z , and obtain upper bounds which are valid in this context. The question of whether or not these bounds suffice to determine convergence is postponed until Section 6.2.

A lace $L \in \mathcal{L}_N[0, |\omega|]$ consists of exactly N edges, and hence the factor $\prod_{st \in L} \mathcal{U}_{st}$ occurring in the sum over ω takes on either the value $(-1)^N$ or 0. The nonzero value will be attained for those walks ω such that $\omega(s) = \omega(t)$ for each $st \in L$.

Consider the case $N = 1$. There is a unique lace in $\mathcal{L}_1[0, |\omega|]$, namely $L_1 = \{0, |\omega|\}$. The first product in (5.4.1) is nonzero, in fact -1 , for precisely those walks which end at the origin. Thus $\hat{\Pi}_z^{(1)}(0, x) = 0$ if $x \neq 0$. All edges other than $\{0, |\omega|\}$ are compatible with L_1 , and hence the second product in (5.4.1) is nonzero only for those walks which have no other self-intersections. The sum over walks in $\hat{\Pi}_z^{(1)}(0, x)$ is thus the sum over all self-avoiding loops (walks which begin and end at the origin and otherwise are self-avoiding, consisting of at least two steps). Let $\mathcal{U}$ denote the set of

all self-avoiding loops. Then, introducing a diagrammatic notation, we can write

$$\Pi_z^{(1)}(0, x) = \delta_{0,x} \sum_{\omega \in \mathcal{U}} z^{|\omega|} \equiv \delta_{0,x} 0 \bigcirc.$$

We can turn the above into a bound on $\hat{\Pi}_z^{(1)}(k)$ by simply noting that

$$\hat{\Pi}_z^{(1)}(k) = \sum_{\omega \in \mathcal{U}} z^{|\omega|} = \sum_{x \in \Omega} z G_z(0, x) \quad (5.4.2)$$

and hence

$$|\hat{\Pi}_z^{(1)}(k)| \leq |z| \Omega \sup_{x \neq 0} G_{|z|}(0, x). \quad (5.4.3)$$

Writing $\|\cdot\|_{\infty}$ for the x -space supremum norm

$$\|f\|_{\infty} = \sup_{x \in \mathbb{Z}^d} |f(x)|, \quad (5.4.4)$$

and introducing

$$\begin{aligned} H_z(x, y) &= G_z(x, y) - \delta_{x,y} \\ &= \begin{cases} G_z(x, y) & x \neq y \\ 0 & x = y, \end{cases} \end{aligned} \quad (5.4.5)$$

(5.4.2) can be rewritten as

$$|\hat{\Pi}_z^{(1)}(k)| \leq |z| \Omega \|H_{|z|}\|_{\infty}. \quad (5.4.6)$$

(Actually H_z is a function of two variables, and in writing norms of H_z we mean norms of the function $H_z(0, \cdot)$ of a single variable.)

For $N = 2$, we proceed as follows. Laces in $\mathcal{L}_2[0, |\omega|]$ are in a one-one correspondence with pairs of times s_2, t_1 with $0 < s_2 < t_1 < |\omega|$, as can be seen from Figure 5.5. For such a lace L , a walk with $\prod_{st \in L} \mathcal{U}_{st}(\omega) \neq 0$ breaks up in a natural way into three subwalks. Letting x denote the endpoint of ω , these three subwalks are of the form $\omega^{(1)}: 0 \rightarrow x$, $\omega^{(2)}: x \rightarrow 0$, $\omega^{(3)}: 0 \rightarrow x$, of respective lengths $s_2, t_1 - s_2, |\omega| - t_1$. We can split the factor $z^{|\omega|}$ into three corresponding factors

$$z^{|\omega|} = \prod_{i=1}^3 z^{|\omega^{(i)}|}. \quad (5.4.7)$$

Each of the subwalks consists of at least one step. The factor $\prod_{s't' \in \mathcal{C}(L)} (1 + \mathcal{U}_{s't'})$ ensures that each of these three subwalks is self-avoiding, since any

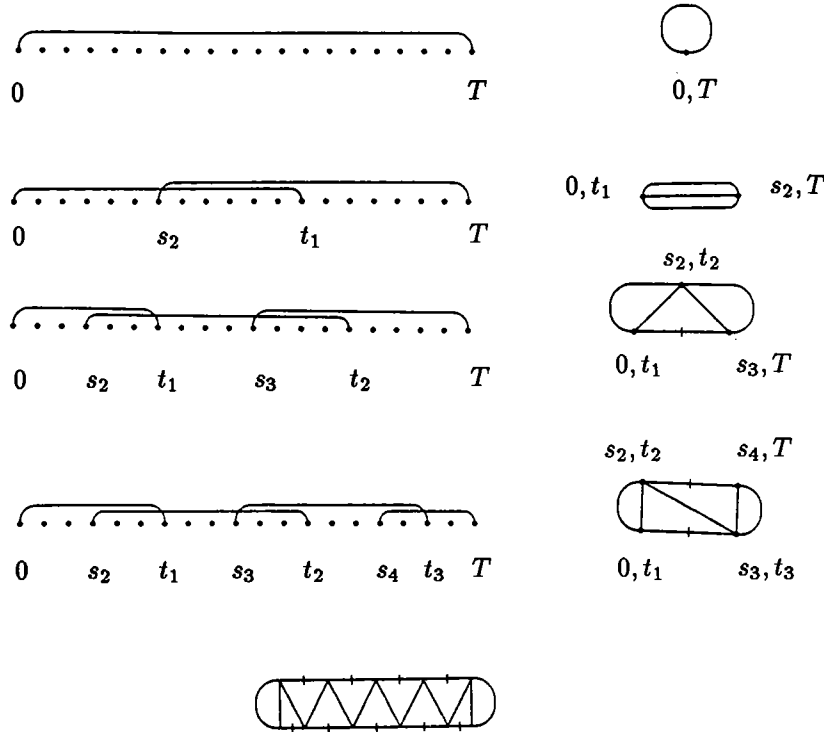


Figure 5.5: The left column shows the general form of laces consisting of 1, 2, 3 or 4 edges. The lace edges are denoted s_i, t_i , $1 \leq i \leq N$, with $s_1 = 0$ and $t_N = T$. The right column shows the self-intersections required for a walk ω with $\prod_{s,t \in L} \mathcal{U}_{s,t}(\omega) \neq 0$. For $\prod_{s,t \in L} (1 + \mathcal{U}_{s,t}) \neq 0$, each of the $2N - 1$ subwalks must be a self-avoiding walk, and in addition there must be mutual avoidance between some (but not all) of the subwalks. The number of loops in a diagram is equal to the number of edges in the corresponding lace. The lines which are slashed correspond to subwalks which may consist of zero steps, but the others correspond to subwalks consisting of at least one step. The eleven-loop diagram is depicted at the bottom.

bond lying entirely in one of the intervals $[0, s_2]$, $[s_2, t_1]$, $[t_1, |\omega|]$ is compatible with L . (In particular, $x \neq 0$.) This same factor also provides an avoidance interaction between the three subwalks, since there are bonds in $\mathcal{C}(L)$ which link distinct subwalks. This interaction between distinct subwalks would be awkward to take into account exactly, but in an upper bound it can be disregarded since $1 + \mathcal{U}_{s,t} \leq 1$. Therefore

$$|\Pi_z^{(2)}(0, x)| \leq H_{|z|}(0, x)^3. \quad (5.4.8)$$

We now show how (5.4.8) can be used to estimate $\hat{\Pi}_z^{(2)}(k)$ and its second derivative with respect to a component k_μ of k . Writing ∂_μ^u for the u -th order partial derivative with respect to k_μ , and using the definition of the Fourier transform,

$$\begin{aligned} |\partial_\mu^u \hat{\Pi}_z^{(2)}(k)| &\leq \sum_x |x_\mu|^u |\Pi_z^{(2)}(0, x)| \leq \sum_x |x_\mu|^u H_{|z|}(0, x)^3 \\ &\leq \| |x_\mu|^u H_{|z|}(0, x) \|_\infty \|H_{|z|}\|_2^2. \end{aligned} \quad (5.4.9)$$

Here $\|\cdot\|_2$ denotes the x -space L^2 -norm

$$\|f\|_2 = \left[\sum_{x \in \mathbb{Z}^d} |f(x)|^2 \right]^{1/2} \quad (5.4.10)$$

In terms of the bubble diagram (1.5.4), the right side of (5.4.9) is equal to

$$\| |x_\mu|^u H_{|z|}(0, x) \|_\infty [\mathcal{B}(|z|) - 1]. \quad (5.4.11)$$

This bound provides an indication of the critical nature of $d = 4$, in the following way. Assuming that the infrared bound $\eta \geq 0$ is indeed valid, then $\mathcal{B}(z_c)$ is finite for $d > 4$. The infrared bound also implies that the critical two-point function decays at least as fast as $|x|^{2-d}$, so that for $d > 4$ (5.4.11) will be finite at the critical point for $u \leq 2$, and hence so will $\partial_\mu^u \hat{\Pi}_{z_c}^{(2)}(k)$. For models with a suitable weak interaction, such as the nearest-neighbour model in sufficiently high dimensions, or a sufficiently “spread-out” model above four dimensions (see Chapter 6), the quantity $\mathcal{B}(z_c) - 1$ will be not only finite, but arbitrarily small, and will be the small parameter responsible for convergence of the expansion. For this reason the distinction between H_z and G_z will be crucial.

We wish to obtain bounds on $\hat{\Pi}_z^{(N)}(k)$ and some of its derivatives in an analogous way for higher values of N . To state these bounds we first need to introduce the following definition.

In the product over j on the left side, factors are to be taken with decreasing index from left to right.

Proof. Fix $k \in \{0, 1, 2, \dots, 2M\}$. Using $\|\mathcal{A}\|_{p,r}$ to denote the norm of an operator $\mathcal{A} : \ell^p \rightarrow \ell^r$, the left side of (5.4.16) can be bounded above by

$$\prod_{j>k} \|\mathcal{H}_{2j}\mathcal{M}_{2j}\|_{\infty,\infty} \|\mathcal{H}_{2k}\mathcal{M}_{2k}\|_{2,\infty} \prod_{i<k} \|\mathcal{H}_{2i}\mathcal{M}_{2i}\|_{2,2} \|f_0\|_p \quad (5.4.17)$$

where $p = 2$ if $k > 0$ and $p = \infty$ if $k = 0$. (Also the norm of $\mathcal{H}_0\mathcal{M}_0$ should be omitted if $k = 0$.) The desired result then follows from the inequalities

$$\begin{aligned} \|\mathcal{H}_{2j}\mathcal{M}_{2j}\|_{\infty,\infty} &\leq \|\mathcal{H}_{2j}\|_{2,\infty} \|\mathcal{M}_{2j}\|_{\infty,2} \leq \|f_{2j}\|_2 \|f_{2j-1}\|_2 \\ \|\mathcal{H}_{2j}\mathcal{M}_{2j}\|_{2,2} &\leq \|\mathcal{H}_{2j}\|_{1,2} \|\mathcal{M}_{2j}\|_{2,1} \leq \|f_{2j}\|_2 \|f_{2j-1}\|_2 \\ \|\mathcal{H}_{2k}\mathcal{M}_{2k}\|_{2,\infty} &\leq \|\mathcal{H}_{2k}\|_{2,\infty} \|\mathcal{M}_{2j}\|_{2,2} \leq \|f_{2k}\|_2 \|f_{2k-1}\|_\infty \\ \|\mathcal{H}_{2k}\mathcal{M}_{2k}\|_{2,\infty} &\leq \|\mathcal{H}_{2k}\|_{1,\infty} \|\mathcal{M}_{2j}\|_{2,1} \leq \|f_{2k}\|_\infty \|f_{2k-1}\|_2, \end{aligned}$$

where the right hand inequalities follow from the Hölder and Young inequalities. (The latter states that $\|g * h\|_s \leq \|g\|_r \|h\|_p$ for $1 \leq p, r, s \leq \infty$ satisfying $p^{-1} + r^{-1} = 1 + s^{-1}$.) $\square$

In the next theorem, Lemma 5.4.3 (in conjunction with Remark 1 below Theorem 5.4.2) is used to bound various derivatives of $\hat{\Pi}_z(k)$. For finite memory, the bounds can be improved as in Remark 2 below Theorem 5.4.2.

Theorem 5.4.4 For any $v \geq 0$,

$$|\partial_z^v \hat{\Pi}_z^{(1)}(k)| \leq \Omega \|\partial_z^v |_{z=|z|} [z H_z]\|_\infty. \quad (5.4.18)$$

For any integer $N \geq 2$,

$$|\partial_z \hat{\Pi}_z^{(N)}(k)| \leq (2N-1) \|\partial_z |_{z=|z|} H_z\|_\infty \|H_{|z|}\|_2^N \|G_{|z|}\|_2^{N-2}. \quad (5.4.19)$$

For any integers $v = 0, 1$, $u \geq 0$ and $N \geq 2$,

$$\begin{aligned} |\partial_z^v \partial_\mu^u \hat{\Pi}_z^{(N)}(k)| &\leq (2N-1)^v \left[\frac{N+1}{2} \right]^u \|x_\mu^u H_{|z|}\|_\infty \|\partial_z^v |_{z=|z|} H_z\|_2 \\ &\quad \times \|H_{|z|}\|_2^{N-1} \|G_{|z|}\|_2^{N-2}. \end{aligned} \quad (5.4.20)$$

For any $N \geq 2$ and for any positive p ,

$$\begin{aligned} 0 &\leq \hat{\Pi}_p^{(N)}(0) - \hat{\Pi}_p^{(N)}(k) \\ &\leq \frac{1}{d} \sum_{\mu=1}^d (1 - \cos k_\mu) \left[\frac{N+1}{2} \right]^2 \| |x|^2 H_p \|_\infty \|H_p\|_2^N \|G_p\|_2^{N-2}. \end{aligned} \quad (5.4.21)$$

For $N = 1$, $\hat{\Pi}_z^{(1)}(0) - \hat{\Pi}_z^{(1)}(k) = 0$, and hence $\partial_\mu^u \hat{\Pi}_z^{(1)}(k) = 0$ for all $u \geq 1$.

Proof. The bound (5.4.18) follows immediately from (5.4.2). For (5.4.19), applying ∂_z to $z^{|\omega|}$ brings down a factor $|\omega|$. Considering ω to consist of $2N-1$ subwalks ω_j , we have

$$|\omega| = \sum_{j=0}^{2N-1} |\omega_j|.$$

Now we bound the diagram in which the j -th subwalk is weighted by the factor $|\omega_j|$ using Lemma 5.4.3, with the infinity norm on the j -th subwalk. This gives (5.4.19).

For (5.4.20) the k -derivative gives rise to an additional factor $|x_\mu^u|$. For the factor $|\omega|$ we proceed as for (5.4.19), obtaining a sum of terms in which one subwalk is weighted by a factor $|\omega_j|$. Then for the factor $|x_\mu^u|$ we choose a sequence of subwalks (depending on j) connecting 0 and x spatially, without using the already weighted j -th subwalk. In general no more than $(N+1)/2$ subwalks will be needed. We then write x_μ as the sum of displacements along the subwalks, and use the inequality

$$|a_1 + \dots + a_m|^u \leq m^{u-1} (|a_1|^u + \dots + |a_m|^u)$$

to obtain a sum of diagrams where in each term one subwalk is weighted by $|\omega_j|$ and another by $|x_\mu^u|$. Then Lemma 5.4.3 is applied.

To prove (5.4.21), we note that the first inequality follows from the fact that by symmetry and (5.2.16),

$$\hat{\Pi}_p^{(N)}(0) - \hat{\Pi}_p^{(N)}(k) = \sum_x \Pi_p^{(N)}(0, x) [1 - \cos k \cdot x] \geq 0.$$

For the upper bound we write

$$\hat{\Pi}_p^{(N)}(0) - \hat{\Pi}_p^{(N)}(k) = \sum_x \Pi_p^{(N)}(0, x) (1 - e^{ik \cdot x}), \quad (5.4.22)$$

and write the last factor on the right as a telescoping sum

$$1 - e^{ik \cdot x} = \sum_{\mu=1}^d (1 - e^{ik_\mu x_\mu}) \prod_{\nu < \mu} e^{ik_\nu x_\nu}. \quad (5.4.23)$$

Inserting (5.4.23) into (5.4.22) and using symmetry gives

$$\hat{\Pi}_p^{(N)}(0) - \hat{\Pi}_p^{(N)}(k) = \sum_x \Pi_p^{(N)}(0, x) \sum_{\mu=1}^d (1 - \cos k_\mu x_\mu) \prod_{\nu < \mu} \cos k_\nu x_\nu. \quad (5.4.24)$$

Using the fact that for any real y and positive integer m ,

$$0 \leq 1 - \cos my \leq m^2(1 - \cos y), \quad (5.4.25)$$

and again using symmetry, the right side of (5.4.24) can be bounded above by

$$\frac{1}{d} \sum_{\mu=1}^d (1 - \cos k_\mu) d \sum_x x_\mu^2 \Pi_p^{(N)}(0, x). \quad (5.4.26)$$

The bound (5.4.21) then follows, since it was shown in the proof of (5.4.20) (with $v = 0$) that the sum over x on the right side of the above equation is bounded above by the right side of (5.4.20).

Finally, the last statement of the theorem follows immediately from the first equality of (5.4.2). $\square$

We end this section with a lemma which will be used in Section 6.8 to bound $c_n(0, x)$. In the statement of the lemma, $c_{n,\sigma}(0, x)$ denotes the number of n -step memory- σ walks from 0 to x .

Lemma 5.4.5 *For memories $\sigma < \tau \leq \infty$,*

$$\begin{aligned} |\hat{\Pi}_z(k; \sigma) - \hat{\Pi}_z(k; \tau)| &\leq 2 \left\| \sum_{n=\sigma/6}^{\infty} c_{n,\sigma}(0, x) |z|^n \right\|_{\infty} \\ &\times \left[|z| \Omega + \sum_{N=2}^{\infty} (2N-1) \|H|_z|(0, \cdot; \sigma)\|_2^N \|G|_z|(0, \cdot; \sigma)\|_2^{N-2} \right]. \end{aligned} \quad (5.4.27)$$

Proof. By Definition 5.2.1, $\mathcal{L}_\sigma[a, b] \subset \mathcal{L}_\tau[a, b]$, and for $L \in \mathcal{L}_\sigma[a, b]$, we have $\mathcal{C}_\sigma(L) \subset \mathcal{C}_\tau(L)$. The latter inclusion is often strict, but if L has no bond whose length exceeds $\sigma/2$ then $\mathcal{C}_\tau(L)$ can contain no bond whose length exceeds σ , and hence $\mathcal{C}_\sigma(L) = \mathcal{C}_\tau(L)$. Therefore in the difference on the left side of (5.4.27) there will be an exact cancellation of terms arising from laces having all bonds of length less than or equal to $\sigma/2$. Temporarily writing $\tilde{\Pi}_m$ for the contribution to $\hat{\Pi}_z(k; m)$ due to laces having at least one bond of length exceeding $\sigma/2$, we have

$$|\hat{\Pi}_z(k; \sigma) - \hat{\Pi}_z(k; \tau)| \leq |\tilde{\Pi}_\sigma| + |\tilde{\Pi}_\tau|. \quad (5.4.28)$$

A typical bond in a lace spans either two or three subwalks; see Figure 5.5. For a lace bond of length at least $\sigma/2$, at least one of the subwalks must consist of at least $\sigma/6$ steps. Now we bound the right side of (5.4.28) as in (5.4.20) (with $u = v = 0$) and (5.4.3), using the L^∞ norm on a subwalk consisting of at least $\sigma/6$ steps. The factor $2N - 1$ arises from the

fact that any one of the $2N - 1$ subwalks may be the only subwalk of at least $\sigma/6$ steps. Also, the bound on $\tilde{\Pi}_\tau$ can be written in terms of memory σ rather than τ , since $c_{n,\tau}(0, x) \leq c_{n,\sigma}(0, x)$.

Finally we observe that in the L^2 norms on the right side, the generating functions H and G can be truncated at the term of order z^σ by the second Remark under Theorem 5.4.2. $\square$

5.5 Other models

The lace expansion has been adapted to study lattice trees and animals, and percolation, and provides a way of proving mean-field critical behaviour above eight and six dimensions respectively. For these models it is not the finiteness of the bubble diagram which leads to mean-field behaviour, but rather the finiteness of the square diagram for lattice trees and animals, and of the triangle diagram for percolation. The lace expansion can be used to prove that these diagrams are finite.

In this section we discuss the lace expansion for lattice trees and animals and for percolation. The material developed here is not required elsewhere in the book. The proof of convergence of the expansion in these contexts is very similar to the proof of convergence of the expansion for the self-avoiding walk (see Section 6.2), and will not be discussed here. Instead we just derive the expansions, and briefly indicate how they can be bounded. In the bounds the square or triangle diagrams play a key role.

5.5.1 Lattice trees and animals

Let Ω be a finite set of sites in $\mathbf{Z}^d$, not containing the origin, which is symmetric with respect to the symmetries of $\mathbf{Z}^d$. We refer to a pair $\{x, y\}$ of sites with $y - x \in \Omega$ as a bond. A *lattice tree* is a connected set of bonds which has no closed loops. Although a tree T is defined as a set of bonds, we write $x \in T$ if x is an endpoint of some bond of T . We denote the number of bonds in a tree T by $|T|$. A *lattice animal* is a connected set of bonds which may contain closed loops. We denote a typical lattice animal by A and the number of bonds in A by $|A|$.

Let t_n denote the number of n -bond trees containing the origin, and a_n the number of n -bond animals containing the origin. It can be shown using subadditivity arguments that both $t_n^{1/n}$ and $a_n^{1/n}$ converge to finite positive limits λ and λ_a as n goes to infinity. The asymptotic behaviour of these quantities is believed to be of the form

$$t_n \sim \text{const.} \lambda^n n^{-\theta+1}, \quad a_n \sim \text{const.} \lambda_a^n n^{-\theta+1} \quad (5.5.1)$$

for some critical exponent θ which is the same for both trees and animals.

Another quantity of interest is the radius of gyration $R(n)$ (for trees) and $R_a(n)$ (for animals). This is defined by

$$R(n)^2 = \frac{\sum_{T \ni 0, |T|=n} \sum_{x \in T} |x - \bar{x}_T|^2}{\sum_{T \ni 0, |T|=n} \sum_{x \in T} 1} \quad (5.5.2)$$

where $\bar{x}_T = (n+1)^{-1} \sum_{x \in T} x$ is the centre of mass of T and $|x|$ is the Euclidean length of x , and similarly for $R_a(n)$. This is believed to behave as

$$R(n) \sim \text{const.} n^\nu, \quad R_a(n) \sim \text{const.} n^\nu \quad (5.5.3)$$

for some critical exponent ν which again is the same for both trees and animals.

The *two-point function* for trees or animals is defined, for sites $x, y \in \mathbb{Z}^d$, by

$$G_z(x, y) = \sum_{T \ni x, y} z^{|T|}, \quad G_z^a(x, y) = \sum_{A \ni x, y} z^{|A|}, \quad (5.5.4)$$

where z is a complex parameter known as the *activity*. We wish to derive an expansion for these two-point functions analogous to the lace expansion for the self-avoiding walk. This can be done using either the inclusion-exclusion or the resummation approach, and here we shall follow the latter.

Using the expansion, the following theorem is proven in Hara and Slade (1992c). Related results for animals appear in Hara and Slade (1990b).

Theorem 5.5.1 *For Ω the set of nearest neighbours of the origin and d sufficiently large, or for $d > 8$ and $\Omega = \{x \neq 0 : \|x\|_\infty \leq L\}$ with L sufficiently large, there are positive constants A and D such that $\theta = 5/2$ and $\nu = 1/4$ in the sense that for any $\epsilon < \min\{1/2, (d-8)/4\}$,*

$$t_n = A\lambda^n n^{-3/2} [1 + O(n^{-\epsilon})]$$

and

$$R_n = Dn^{1/4} [1 + O(n^{-\epsilon})].$$

The expansion for trees

We begin with lattice trees. Given two distinct sites x, y and a tree $T \ni x, y$, the *backbone* $\beta_T(x, y)$ of T is defined to be the unique path, consisting of bonds of T , which joins x to y . Usually x and y are understood and we write simply β_T for $\beta_T(x, y)$. Sites in the backbone are labelled consecutively

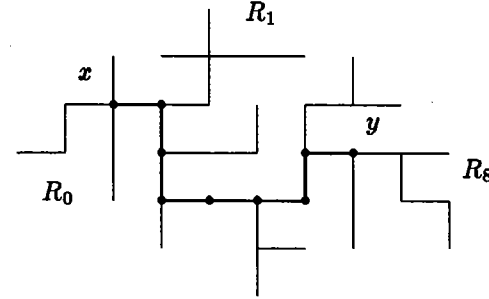


Figure 5.6: Decomposition of a tree T containing sites x and y into its backbone $\beta_T(x, y)$ and ribs $R_0, \dots, R_8$. The vertices of the backbone are indicated by heavy dots.

from x to y , beginning with $\beta_T(0) = x$ and ending at (say) $\beta_T(n) = y$. Removal of the bonds in the backbone disconnects the tree into $n+1$ mutually nonintersecting trees $R_0, \dots, R_n$, which we refer to as *ribs*. This decomposition is shown in Figure 5.6.

Given a set $\mathbf{R} = \{R_0, \dots, R_n\}$ of $n+1$ trees R_i , we define

$$\mathcal{U}_{st}(\mathbf{R}) = \begin{cases} -1 & \text{if } R_s \text{ and } R_t \text{ share a common site} \\ 0 & \text{if } R_s \text{ and } R_t \text{ share no common site.} \end{cases} \quad (5.5.5)$$

Then the two point function can be written

$$G_z(0, x) = \sum_{\omega: 0 \rightarrow x} z^{|\omega|} \left[\prod_{i=0}^{|\omega|-1} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] \prod_{0 \leq s < t \leq |\omega|} (1 + \mathcal{U}_{st}(\mathbf{R})) \quad (5.5.6)$$

where each sum over R_i is a sum over trees containing $\omega(i)$, and $\mathbf{R} = (R_0, \dots, R_{|\omega|})$. The sum over ω is the sum over all ordinary random walks from 0 to x (taking steps in Ω), although walks that are not self-avoiding give zero contribution to (5.5.6).

We use the terminology of Definition 5.2.1 with a change in the definition of connected graph. A graph Γ on $[a, b]$ will now be said to be *connected* if both a and b are endpoints of edges in Γ , and if in addition for each $c \in (a, b)$ there are $s, t \in [a, b]$ such that $s < c < t$ with either (i) $\{s, t\} \in \Gamma$, or (ii) $\{c, t\} \in \Gamma$ and $\{s, c\} \in \Gamma$. This notion of connectedness is less restrictive than that used for the self-avoiding walk, and is better suited for dealing with the interaction between ribs. This new definition of connected graph

leads to a larger set of laces than before, where we still define a *lace* to be a minimally connected graph, i.e., a connected graph for which the removal of any edge would result in a disconnected graph. We modify the prescription associating to each connected graph Γ a unique lace $\mathcal{L}_\Gamma$, to conform with the new notion of connectedness, and now define $\mathcal{L}_\Gamma$ to consist of edges $s_1 t_1, s_2 t_2, \dots$ where

$$\begin{aligned} s_1 &= a, & t_1 &= \max\{t : at \in \Gamma\} \\ t_{i+1} &= \max\{t : st \in \Gamma, s \leq t_i\} \\ s_i &= \min\{s : st_i \in \Gamma\}. \end{aligned}$$

Let

$$K[a, b] = \prod_{a \leq s < t \leq b} (1 + \mathcal{U}_{st}) \quad (5.5.7)$$

and

$$J[0, a] = \sum_{L \in \mathcal{L}[0, a]} \prod_{st \in L} \mathcal{U}_{st} \prod_{s't' \in \mathcal{C}(L)} (1 + \mathcal{U}_{s't'}). \quad (5.5.8)$$

Then following the proof of Lemma 5.2.2 with the modified definition of connectedness gives for $b \geq 1$,

$$K[0, b] = K[1, b] + \sum_{a=1}^{b-1} J[0, a] K[a+1, b] + J[0, b]. \quad (5.5.9)$$

(The middle term on the right side is taken to be 0 if $b = 1$.) Substitution of (5.5.9) into (5.5.6) results in

$$\begin{aligned} G_z(0, x) &= \sum_{R_0 \ni 0} z^{|R_0|} \delta_{0,x} + \sum_{\substack{\omega : 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} \left[\prod_{i=0}^{|\omega|} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] K[1, |\omega|] \\ &+ \sum_{\substack{\omega : 0 \rightarrow x \\ |\omega| \geq 2}} z^{|\omega|} \left[\prod_{i=0}^{|\omega|} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] \sum_{a=1}^{|\omega|-1} J[0, a] K[a+1, |\omega|] \\ &+ \sum_{\substack{\omega : 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} \left[\prod_{i=0}^{|\omega|} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] J[0, |\omega|]. \end{aligned} \quad (5.5.10)$$

The first term on the right hand side is due to the contribution to (5.5.6) from the trivial zero-step walk, and the other terms are due to the walks ω with $|\omega| \geq 1$.

5.5. OTHER MODELS

Denoting $G_z(0, 0)$ by

$$g_z = G_z(0, 0) = \sum_{T \ni 0} z^{|T|} \quad (5.5.11)$$

and writing

$$\Pi_z(0, x) = \sum_{\substack{\omega : 0 \rightarrow x \\ |\omega| \geq 1}} z^{|\omega|} \left[\prod_{i=0}^{|\omega|} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] J[0, |\omega|], \quad (5.5.12)$$

the first and last terms on the right side of (5.5.10) can be written as $g_z \delta_{0,x}$ and $\Pi_z(0, x)$ respectively. The second term on the right side of (5.5.10) is equal to

$$\sum_{R_0 \ni 0} z^{|R_0|} z \sum_{u \in \Omega} G_z(u, x).$$

For the third term on the right side of (5.5.10), we consider ω to be composed of an initial a -step walk ω_1 from 0 to (say) u , followed by a single step to (say) v , and then a final portion (possibly consisting of 0 steps) ω_2 from v to x . The term in question is then equal to

$$\begin{aligned} &z \sum_{(u,v)} \sum_{\substack{\omega_1 : 0 \rightarrow u \\ |\omega_1| \geq 1}} z^{|\omega_1|} \left[\prod_{i=0}^{|\omega_1|} \sum_{R_i \ni \omega_1(i)} z^{|R_i|} \right] J[0, |\omega_1|] \\ &\times \sum_{\substack{\omega_2 : v \rightarrow x \\ |\omega_2| \geq 0}} z^{|\omega_2|} \left[\prod_{i=0}^{|\omega_2|} \sum_{R_i \ni \omega_2(i)} z^{|R_i|} \right] K[0, |\omega_2|] \\ &= z \sum_{(u,v)} \Pi_z(0, u) G_z(v, x), \end{aligned}$$

where the sum over (u, v) denotes the sum over all directed steps with $v - u \in \Omega$. Summarizing, (5.5.10) can be rewritten as

$$G_z(0, x) = \delta_{0,x} g_z + \Pi_z(0, x) + z g_z \sum_{(0,u)} G_z(u, x) + z \sum_{(u,v)} \Pi_z(0, u) G_z(v, x). \quad (5.5.13)$$

This is the expansion for trees.

We can write this more compactly by writing $D(x)$ for Ω^{-1} times the indicator function of the set Ω , and defining

$$h_z(x) = \delta_{0,x} g_z + \Pi_z(0, x). \quad (5.5.14)$$

Then (5.5.13) becomes

$$G_z(0, x) = h_z(x) + h_z * z\Omega D * G_z(x). \quad (5.5.15)$$

The expansion for animals

For lattice animals the derivation of the expansion requires some modification due to the fact that for an animal containing sites x and y there is in general not a unique path in the animal from x to y . To describe this modification, some definitions are needed.

A lattice animal A containing x and y is said to have a *double connection* from x to y if there are two *disjoint* (i.e. sharing no common bond) self-avoiding walks in A between x and y or if $x = y$. A bond $\{u, v\}$ in A is called *pivotal* for the connection from x to y if its removal would disconnect the animal into two connected components with x in one connected component and y in the other. There is a natural order to the set of pivotal bonds for the connection from x to y , and each pivotal bond is ordered in a natural way, as follows. The *first* pivotal bond for the connection from x to y (assuming there is at least one) is the pivotal bond for which there is a double connection between one endpoint of the pivotal bond and x . The endpoint for which there is a double connection to x is then the *first* endpoint of the first pivotal bond. To determine the second pivotal bond, the role of x is then played by the second endpoint of the first pivotal bond, and so on.

Given two sites x, y and an animal A containing x and y , the *backbone* of A is now defined to be the set of pivotal bonds for the connection from x to y . In general this backbone is not connected. The *ribs* of A are the connected components which remain after the removal of the backbone from A . An example is depicted in Figure 5.7. The set of all animals having a double connection between x and y is denoted $\mathcal{D}_{x,y}$, and we write

$$g_z^a(x, y) = \sum_{D \in \mathcal{D}_{x,y}} z^{|D|}. \quad (5.5.16)$$

In particular $\mathcal{D}_{x,x}$ is the set of all animals containing x . Let B be an arbitrary finite ordered set of directed bonds:

$$B = ((u_1, v_1), \dots, (u_{|B|}, v_{|B|})).$$

Let $v_0 = 0$ and $u_{|B|+1} = x$. Then

$$G_z^a(0, x) = \sum_{B: |B| \geq 0} z^{|B|} \left[\prod_{i=0}^{|B|} \sum_{D_i \in \mathcal{D}_{v_i, v_{i+1}}} z^{|D_i|} \right] K[0, |B|],$$

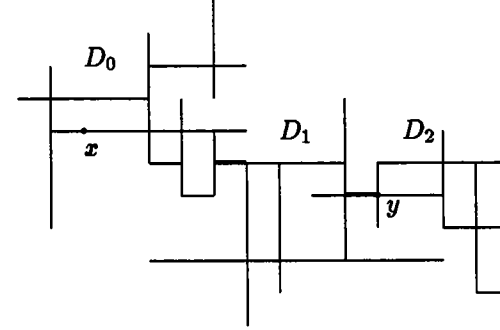


Figure 5.7: Decomposition of a lattice animal A containing x and y into backbone and ribs. The backbone, consisting of two bonds, is drawn in bold lines.

where now in the definition of $K[0, |B|]$ in (5.5.7)

$$u_{st} = \begin{cases} -1 & \text{if } D_s \text{ and } D_t \text{ share a common site} \\ 0 & \text{if } D_s \text{ and } D_t \text{ have no common site.} \end{cases}$$

Define

$$\Pi_z^a(0, y) = \sum_{B: |B| \geq 1} z^{|B|} \left[\prod_{i=0}^{|B|} \sum_{D_i \in \mathcal{D}_{v_i, v_{i+1}}} z^{|D_i|} \right] J[0, |B|] \quad (5.5.17)$$

and

$$h_z^a(x) = g_z^a(0, x) + \Pi_z^a(0, x). \quad (5.5.18)$$

A calculation similar to that used to derive (5.5.13), using (5.5.9), gives

$$G_z^a(0, x) = h_z^a(x) + h_z^a * z\Omega D * G_z^a(x). \quad (5.5.19)$$

Bounds on the expansion for trees

For proving convergence of the expansions for trees, bounds are required on $\hat{\Pi}_z(k)$ and $\partial_\mu^2 \hat{\Pi}_z(k)$. In this section we indicate how appropriate bounds can be obtained on these quantities. The procedure for obtaining bounds for lattice animals is similar but more involved, and will not be discussed here.

We denote by $\mathcal{L}_N[0, a]$ the set of laces in $\mathcal{L}[0, a]$ consisting of exactly N edges, and write

$$J_N[0, a] = \sum_{L \in \mathcal{L}_N[0, a]} \prod_{st \in L} \mathcal{U}_{st} \prod_{s't' \in C(L)} (1 + \mathcal{U}_{s't'}) \quad (5.5.20)$$

and

$$\Pi_z^{(N)}(0, x) = (-1)^N \sum_{\substack{\omega: 0 \rightarrow x, \\ |\omega| \geq 1}} z^{|\omega|} \left[\prod_{i=0}^{|\omega|-1} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] J_N[0, |\omega|]. \quad (5.5.21)$$

Then from (5.5.12) and (5.5.8) we have

$$\Pi_z(0, x) = \sum_{N=1}^{\infty} (-1)^N \Pi_z^{(N)}(0, x). \quad (5.5.22)$$

For a nonzero contribution to $\Pi_z^{(N)}(0, x)$, the factor $\prod_{st \in L} \mathcal{U}_{st}$ in J_N enforces intersections between the ribs R_s and R_t . This leads to bounds in which the contribution to $\Pi_z^{(N)}(0, x)$ from the N -edge laces can be bounded above by an N -loop diagram. We illustrate this in detail only for the simplest cases $N = 1, 2$. We also discuss the manner of bounding diagrams only for these two simplest cases. For more details the reader is referred to Hara and Slade (1990b).

To bound the term $\Pi_z^{(1)}(0, x)$ we proceed as follows. There is a unique lace consisting of a single edge, so by definition

$$\begin{aligned} \Pi_z^{(1)}(0, x) &= - \sum_{\substack{\omega: 0 \rightarrow x, \\ |\omega| \geq 1}} z^{|\omega|} \left[\prod_{i=0}^{|\omega|-1} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] \mathcal{U}_{0, |\omega|} \\ &\times \prod_{\substack{0 \leq s < t \leq |\omega| \\ (s, t) \neq (0, |\omega|)}} (1 + \mathcal{U}_{st}). \end{aligned} \quad (5.5.23)$$

The factor $\mathcal{U}_{0, |\omega|}$ gives a nonzero contribution only if R_0 and $R_{|\omega|}$ intersect, and the final product in (5.5.23) disallows any further rib intersections. We first consider the case $x \neq 0$. Relaxing the latter restriction somewhat and overcounting an enforcement of the former gives the upper bound

$$|\Pi_z^{(1)}(0, x)| \leq \sum_v \sum_{\substack{\omega: 0 \rightarrow x, \\ |\omega| \geq 1}} |z|^{|\omega|} \sum_{R_0 \ni 0, v} |z|^{|R_0|} \sum_{R_{|\omega|} \ni x, v} |z|^{|R_{|\omega|}|}$$

5.5. OTHER MODELS

$$\times \left[\prod_{i=1}^{|\omega|-1} \sum_{R_i \ni \omega(i), \neq 0, x} |z|^{|R_i|} \right] \prod_{1 \leq s < t \leq |\omega|-1} (1 + \mathcal{U}_{st}).$$

Now

$$\sum_{R_0 \ni 0, v} |z|^{|R_0|} = G_{|z|}(0, v)$$

and

$$\sum_{R_{|\omega|} \ni x, v} |z|^{|R_{|\omega|}|} = G_{|z|}(x, v).$$

Since

$$\sum_{\substack{\omega: 0 \rightarrow x, \\ |\omega| \geq 1}} |z|^{|\omega|} \left[\prod_{i=1}^{|\omega|-1} \sum_{R_i \ni \omega(i), \neq 0, x} |z|^{|R_i|} \right] \prod_{1 \leq s < t \leq |\omega|-1} (1 + \mathcal{U}_{st}) \leq G_{|z|}(0, x),$$

for $x \neq 0$ we have

$$|\Pi_z^{(1)}(0, x)| \leq \sum_v G_{|z|}(0, x) G_{|z|}(x, v) G_{|z|}(v, 0). \quad (5.5.24)$$

When $x = 0$ in (5.5.23), we can argue similarly that

$$|\Pi_z^{(1)}(0, 0)| \leq \sum_{|e|=1} G_{|z|}(0, e) G_{|z|}(e, 0), \quad (5.5.25)$$

by decoupling the last step of ω from its preceding steps. Since $G_{|z|}(0, 0) \geq 1$, we can now combine (5.5.24) and (5.5.25) to obtain

$$|\Pi_z^{(1)}(0, x)| \leq \sum_v \left[G_{|z|}(0, v) G_{|z|}(v, x) G_{|z|}(x, 0) - \delta_{0, v} \delta_{v, x} g_{|z|}^3 \right]. \quad (5.5.26)$$

Thus $\hat{\Pi}_z^{(1)}(k)$ is bounded by the triangle diagram minus the trivial contribution arising when all three vertices are the origin. For $\partial_\mu^2 \hat{\Pi}_z^{(1)}(k)$, the corresponding bound is the triangle diagram with one line weighted by x_μ^2 .

A similar strategy can be used to bound $\Pi_z^{(N)}(0, x)$ for $N \geq 2$. The situation for $N = 2$ is shown in Figure 5.8. For a lace $L = (0t_1, s_2|\omega|)$ consisting of exactly two edges, there are two generic configurations possible with $\prod_{st \in L} \mathcal{U}_{st} \neq 0$, one for the case $s_2 < t_1$ and the other for $s_2 = t_1$. We illustrate the basic idea for the latter case. The contribution to $\Pi_z^{(2)}$ due to

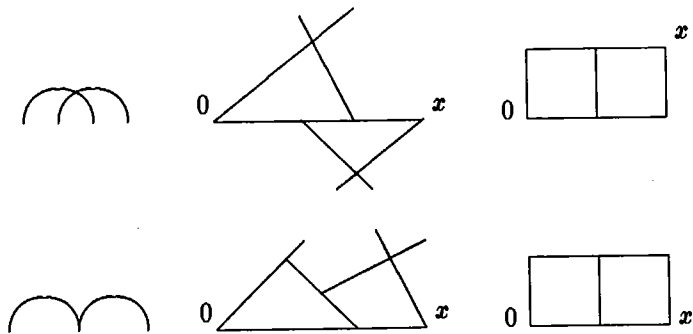


Figure 5.8: The two generic laces consisting of two bonds, schematic diagrams showing the corresponding rib intersections for a nonzero contribution to $\Pi_z^{(2)}(0, x)$, and Feynman diagrams bounding the corresponding contributions to $\Pi_z^{(2)}(0, x)$. In evaluating the diagrams, sums over vertices are constrained to disallow the coincidence of all vertices on any loop.

laces with $s_2 = t_1$ can be written

$$\sum_{\substack{\omega: 0 \rightarrow x \\ |\omega| \geq 2}} z^{|\omega|} \left[\prod_{i=0}^{|\omega|-1} \sum_{R_i \ni \omega(i)} z^{|R_i|} \right] \sum_{s=1}^{|\omega|-1} \mathcal{U}_{0s} \mathcal{U}_{s|\omega|} \prod_{s' t' \in \mathcal{C}(\{0, s\}, \{s, |\omega|\})} (1 + \mathcal{U}_{s' t'}). \quad (5.5.27)$$

We bound this by replacing z by $|z|$, and using $1 + \mathcal{U}_{s' t'} \leq 1$ for $s' < s < t'$ to decouple the interaction. For $\mathcal{U}_{0s} \neq 0$, R_0 and R_s must intersect. Let y be a site where R_0 and R_s intersect, and let $\beta_{R_s}(\omega(s), y)$ be the backbone of R_s . For $\mathcal{U}_{s|\omega|} \neq 0$, $R_{|\omega|}$ must intersect R_s , and hence there must be a rib emanating from a site on the backbone $\beta_{R_s}(\omega(s), y)$ which intersects $R_{|\omega|}$ — see Figure 5.8. Now by arguing in a similar fashion to the case $N = 1$, (5.5.27) can be bounded above by the second Feynman diagram depicted in Figure 5.8. The contribution from the other type of lace is bounded by the first Feynman diagram of Figure 5.8.

In general, $|\hat{\Pi}_z^{(N)}(k)|$ is bounded above by a sum of 2^{N-1} “ladder” diagrams, each containing exactly N non-trivial loops. The basic operation used in bounding these diagrams is repeatedly to apply the simple inequality

$$\left| \sum_x f(x) g(x) \right| \leq \sup_x |f(x)| \sum_y |g(y)|. \quad (5.5.28)$$

For the case $N = 2$, the argument goes as follows. Let

$$\begin{aligned} \bar{T}_z &= \sup_{w \in \mathbb{Z}^d} \left[\sum_{x, y} G_z(0, x) G_z(x, y) G_z(y, w) - \delta_{0, w} (g_z)^3 \right] \\ \bar{W}_z &= \sup_{w \in \mathbb{Z}^d} \sum_{x, y} |x|^2 G_z(0, x) G_z(x, y) G_z(y, w) \\ S_z &= \sum_{w, x, y} G_z(0, w) G_z(w, x) G_z(x, y) G_z(y, 0) - (g_z)^4. \end{aligned}$$

Then we have

$$|\hat{\Pi}_z^{(2)}(k)| \leq 2 \bar{T}_{|z|} S_{|z|}.$$

Also, using the inequality $|\sum_{i=1}^n a_i|^2 \leq n \sum_{i=1}^n |a_i|^2$ to replace $|x|^2$ leads to a sum of diagrams in which a single line in each diagram is weighted by the square of the difference of the line endpoints, and we obtain

$$|\nabla_k^2 \hat{\Pi}_z^{(2)}(k)| \leq 13 \bar{W}_{|z|} S_{|z|}.$$

Here the factor $13 = 3^2 + 2^2$ arises from the fact that in the first diagram of Figure 5.8 the $|x|^2$ is distributed over $n = 3$ lines, while in the second $n = 2$.

Higher values of N can be handled in a similar fashion.

5.5.2 Percolation

Grimmett (1989) provides a good introduction to percolation. Here we mention only those aspects of percolation which are relevant for the development of the lace expansion. The expansion can be applied to both bond and site percolation, but for simplicity we restrict attention to bond percolation.

Let Ω be a finite set of sites in $\mathbb{Z}^d$, not containing the origin, which is symmetric with respect to the symmetries of $\mathbb{Z}^d$. We consider independent bond percolation on $\mathbb{Z}^d$, where the bonds are the pairs $\{x, y\}$ of sites with $y - x \in \Omega$. This means that to each bond we associate an independent Bernoulli random variable $n_{\{x, y\}}$ which takes the value one with probability p and the value zero with probability $1 - p$, where p is a parameter in the closed interval $[0, 1]$. If $n_{\{x, y\}} = 1$ then we say that the bond $\{x, y\}$ is *occupied*, and otherwise we say that it is *vacant*. A *configuration* is a realization of the random variables for all bonds. Given a configuration and any two sites x and y , we say that x and y are *connected* if there is a self-avoiding walk from x to y consisting of occupied bonds, or if $x = y$. For $d \geq 2$ this model undergoes a phase transition in the following sense:

There is a critical value $p_c \in (0, 1)$ such that the probability $\theta(p)$ that the origin is connected to infinitely many sites is zero for $p < p_c$ and strictly positive for $p > p_c$.

We denote the indicator function of an event E by $I[E]$ and expectation with respect to the joint distribution of the Bernoulli random variables $n_{\{x,y\}}$ by $\langle \cdot \rangle_p$. Then the two-point function $\tau_p(x, y)$ is defined to be the probability that x and y are connected:

$$\tau_p(x, y) = \langle I[x \text{ and } y \text{ are connected}] \rangle_p. \quad (5.5.29)$$

For $p < p_c$ the two-point function is known to decay exponentially as $|x - y| \rightarrow \infty$, so that the correlation length

$$\xi(p) = - \left[\lim_{n \rightarrow \infty} \frac{1}{n} \log \tau_p(0, ne_1) \right]^{-1} \quad (5.5.30)$$

is finite and strictly positive. The susceptibility is defined by

$$\chi(p) = \sum_{x \in \mathbb{Z}^d} \tau_p(0, x). \quad (5.5.31)$$

The susceptibility is known to be finite for $p < p_c$ and to diverge as $p \nearrow p_c$.

The following power laws are believed to hold:

$$\chi(p) \sim \text{const.} (p_c - p)^{-\gamma} \text{ as } p \nearrow p_c \quad (5.5.32)$$

$$\theta(p) \sim \text{const.} (p - p_c)^\beta \text{ as } p \searrow p_c \quad (5.5.33)$$

$$\xi(p) \sim \text{const.} (p_c - p)^{-\nu} \text{ as } p \nearrow p_c \quad (5.5.34)$$

for some dimension-dependent critical exponents γ, β, ν . Currently the only rigorous result proving existence of these critical exponents is the following theorem. In the statement of the theorem we as usual write $f(x) \simeq g(x)$ to mean that there are positive constants c_1, c_2 such that $c_1 g(x) \leq f(x) \leq c_2 g(x)$ for x sufficiently close to its limiting value. References to further results in this direction are given in the Notes at the end of the chapter.

Theorem 5.5.2 For Ω the set of nearest neighbours of the origin and d sufficiently large, or for $d > 6$ and $\Omega = \{x \neq 0 : \|x\|_\infty \leq L\}$ with L sufficiently large,

$$\begin{aligned} \chi(p) &\simeq (p_c - p)^{-1} & \text{as } p \nearrow p_c \\ \theta(p) &\simeq (p - p_c)^1 & \text{as } p \searrow p_c \\ \xi(p) &\simeq (p_c - p)^{-1/2} & \text{as } p \nearrow p_c. \end{aligned}$$

It is a short step from the behaviour of $\theta(p)$ given in Theorem 5.5.2 to the conclusion that the percolation probability is zero at the critical point, i.e. that $\theta(p_c) = 0$. This has otherwise been proven only for the nearest-neighbour model in two dimensions; see the Notes for more details.

An important ingredient in the proof of Theorem 5.5.2 is an expansion for the two-point function $\tau_p(0, x)$, valid for $p < p_c$, which is analogous to the expansions for the self-avoiding walk and lattice trees and animals. Our purpose in the remainder of this section is derive the expansion, and to indicate how it can be bounded.

The expansion

For the expansion we need to define several concepts.

Definition 5.5.3 (a) A bond is an unordered pair of distinct sites $\{x, y\}$ with $y - x \in \Omega$. A directed bond is an ordered pair (x, y) of distinct sites with $y - x \in \Omega$. A path from x to y is a self-avoiding walk from x to y , considered to be a set of bonds. Two paths are disjoint if they have no bonds in common (they may have common sites). Given a bond configuration, an occupied path is a path consisting of occupied bonds.

(b) Given a bond configuration, two sites x and y are connected if there is an occupied path from x to y or if $x = y$. We denote by $C(x)$ the random set of sites which are connected to x . Two sites x and y are doubly-connected if there are at least two disjoint occupied paths from x to y or if $x = y$. We denote by $D_c(x)$ the random set of sites which are doubly-connected to x . Given a bond $b = \{u, v\}$ and a bond configuration, we define $C_b(x)$ to be the set of sites which remain connected to x in the new configuration obtained by setting $n_b = 0$.

(c) Given a set of sites $A \subset \mathbb{Z}^d$ and a bond configuration, two sites x and y are connected in A if there is an occupied path from x to y having all of its sites in A (so in particular it is required that $x, y \in A$), or if $x = y \in A$. Two sites x and y are connected through A if they are connected in such a way that every occupied path from x to y has at least one bond with an endpoint in A , or if $x = y \in A$.

(d) The restricted two-point function is defined by

$$\tau_p^A(x, y) = \langle I[x \text{ and } y \text{ are connected in } \mathbb{Z}^d \setminus A] \rangle_p.$$

(e) Given a bond configuration, a bond $\{u, v\}$ (occupied or not) is called pivotal for the connection from x to y if (i) either $x \in C(u)$ and $y \in C(v)$, or $x \in C(v)$ and $y \in C(u)$, and (ii) $y \notin C_{\{u,v\}}(x)$. Similarly a directed bond (u, v) is pivotal for the connection from x to y if $x \in C_{\{u,v\}}(u)$, $y \in$

$C_{\{u,v\}}(v)$ and $y \notin C_{\{u,v\}}(x)$. If x and y are connected then there is a natural order to the set of occupied pivotal bonds for the connection from x to y (assuming there is at least one occupied pivotal bond), and each of these pivotal bonds is directed in a natural way, as follows. The first pivotal bond from x to y is the directed occupied pivotal bond (u, v) such that u is doubly-connected to x . If (u, v) is the first pivotal bond for the connection from x to y , then the second pivotal bond is the first pivotal bond for the connection from v to y , and so on.

The basic idea behind the expansion is similar to that underlying the expansion for lattice animals. Given a configuration in which 0 and x are connected, the connected bond cluster of the origin is a lattice animal containing the sites 0 and x . The occupied pivotal bonds divide the cluster into doubly-connected pieces, as in Figure 5.7. No two of these pieces can share a common site, so there is a kind of “repulsive interaction” between these pieces. However the situation is not as simple as it was for lattice animals, because the pieces interact also when they share a common boundary bond. Rather than try to formalize this interaction into a quantity $\mathcal{U}$ as we did for lattice animals, we shall proceed instead to derive the expansion along the lines of the inclusion-exclusion approach outlined in Section 5.1.

In this approach, we think of the pivotal bonds as corresponding to the steps of a walk. The first thing to do is to extract the contribution due to the zero-step walk, which in the percolation context corresponds to the event that 0 and x are doubly-connected. Thus we have

$$\tau_p(0, x) = \langle I[x \in D_c(0)] \rangle_p + \langle I[x \in C(0), x \notin D_c(0)] \rangle_p. \quad (5.5.35)$$

If 0 is connected to x , but not doubly, then there is a pivotal bond for the connection from 0 to x and hence a first pivotal bond, so that

$$\begin{aligned} \tau_p(0, x) &= \langle I[x \in D_c(0)] \rangle_p \\ &+ \sum_{(u,v)} \langle I[x \in C(0), (u, v) \text{ is the first pivotal bond}] \rangle_p. \end{aligned} \quad (5.5.36)$$

To proceed further, we need a way of writing the last term on the right side as a convolution with τ_p . This is achieved using the next lemma.

For the statement of the lemma, given sites x, y and a set of sites A we introduce two events. Let $E_1(x, y)$ be the event that $y \in D_c(x)$, and let $E_2(x, y; A)$ be the event that x is connected to y through A and there is no pivotal bond for the connection from x to y whose first endpoint is connected to x through A . In particular, $E_2(x, y; A)$ includes the event that x and y are doubly-connected and connected through A . Observe that $E_1(x, y) = E_2(x, y; \mathbb{Z}^d)$.

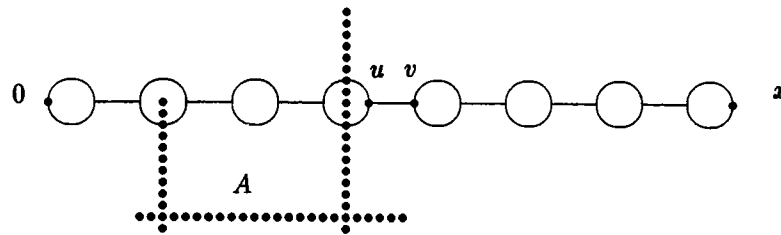


Figure 5.9: The event of Lemma 5.5.4, that $E_2(0, u; A)$ occurs and (u, v) is occupied and pivotal for the connection from 0 to x . The dotted lines represent the sites in A . There is no restriction on intersections between A and $C_{\{u,v\}}(x)$.

Lemma 5.5.4 *Given a nonempty set of sites A and a site u , let $E = E_2(0, u; A)$. Let $p < p_c$. Then*

$$\begin{aligned} &\langle I[E] I[(u, v) \text{ is occupied and pivotal for the connection from 0 to } x] \rangle_p \\ &= p \langle I[E] \tau_p^{C_{\{u,v\}}(0)}(v, x) \rangle_p. \end{aligned} \quad (5.5.37)$$

Proof. The event appearing in the left side of (5.5.37) is depicted in Figure 5.9. The proof is by conditioning on $C_{\{u,v\}}(0)$, which is the connected cluster of the origin which remains after setting $n_{\{u,v\}} = 0$. This cluster is finite with probability one, since $p < p_c$.

We first observe that for the event E under consideration, the event that E occurs and (u, v) is pivotal (for the connection from 0 to x) is independent of the occupation status of the bond (u, v) . Therefore the left side of the identity in the statement of the lemma is equal to

$$p \langle I[E] I[(u, v) \text{ is pivotal for the connection from 0 to } x] \rangle_p. \quad (5.5.38)$$

By conditioning on $C_{\{u,v\}}(0)$, (5.5.38) is equal to

$$p \sum_{S: S \ni 0} \langle I[E \text{ occurs}, (u, v) \text{ is pivotal}, C_{\{u,v\}}(0) = S] \rangle_p, \quad (5.5.39)$$

where the sum is over all finite sets of sites S containing 0.

In (5.5.39), the statement that (u, v) is pivotal can be replaced by the statement that v is connected to x in $\mathbb{Z}^d \setminus S$. This event depends only on the occupation status of the bonds which do not have an endpoint in S . On the other hand, the event E under consideration is determined by the

occupation status of bonds which have an endpoint in $C_{\{u,v\}}(0)$. Similarly, the set $C_{\{u,v\}}(0)$ is *self-determined* in the sense that for a deterministic (finite) set of sites S , the event that $C_{\{u,v\}}(0) = S$ depends on the values of n_b only for bonds b which have one or both endpoints in S . Hence the event that both E occurs and $C_{\{u,v\}}(0) = S$ is independent of the event that v is connected to x in $\mathbf{Z}^d \setminus S$, and therefore (5.5.39) is equal to

$$p \sum_{S: S \ni 0} \langle I[E \text{ occurs and } C_{\{u,v\}}(0) = S] \rangle_p \tau_p^S(v, x). \quad (5.5.40)$$

Bringing the restricted two-point function inside the expectation, replacing the superscript S by $C_{\{u,v\}}(0)$, and performing the sum over S , (5.5.40) is equal to

$$p \langle I[E] \tau_p^{C_{\{u,v\}}(0)}(v, x) \rangle_p. \quad (5.5.41)$$

This completes the proof. $\square$

We now apply the lemma to the second term on the right side of (5.5.36), with $E = E_1(0, u) = E_2(0, u; \mathbf{Z}^d)$. The summand in this term is equal to the probability that 0 is doubly-connected to u and (u, v) is occupied and pivotal for the connection from 0 to x . Hence by the lemma it is equal to

$$p \langle I[u \in D_c(0)] \tau_p^{C_{\{u,v\}}(0)}(v, x) \rangle_p. \quad (5.5.42)$$

To extract a term involving a convolution with τ_p from this quantity, we write

$$\tau_p^{C_{\{u,v\}}(0)}(v, x) = \tau_p(v, x) - [\tau_p(v, x) - \tau_p^{C_{\{u,v\}}(0)}(v, x)]. \quad (5.5.43)$$

Using (5.5.42) and (5.5.43) in (5.5.36), we obtain

$$\begin{aligned} \tau_p(0, x) &= \langle I[x \in D_c(0)] \rangle_p + p \sum_{(u,v)} \langle I[u \in D_c(0)] \rangle_p \tau_p(v, x) \\ &\quad - p \sum_{(u,v)} \langle I[u \in D_c(0)] \{ \tau_p(v, x) - \tau_p^{C_{\{u,v\}}(0)}(v, x) \} \rangle_p. \end{aligned} \quad (5.5.44)$$

The above equation gives the lowest order expansion with remainder. We abbreviate the notation by writing

$$g_p(0, x) = \langle I[x \in D_c(0)] \rangle_p \quad (5.5.45)$$

and

$$R_p^{(0)}(0, x) = p \sum_{(u,v)} \langle I[u \in D_c(0)] \{ \tau_p(v, x) - \tau_p^{C_{\{u,v\}}(0)}(v, x) \} \rangle_p. \quad (5.5.46)$$

5.5. OTHER MODELS

We denote by D the function on $\mathbf{Z}^d$ which takes the value Ω^{-1} at sites in Ω and otherwise is zero. Then (5.5.44) can be rewritten as

$$\tau_p(0, x) = g_p(0, x) + g_p * p\Omega D * \tau_p(x) - R_p^{(0)}(0, x). \quad (5.5.47)$$

To proceed further, we will use the following lemma to expand the remainder term $R_p^{(0)}(0, x)$. For the statement of the lemma, we write $I_2(v, x; A)$ for the indicator function of the event $E_2(v, x; A)$.

Lemma 5.5.5 *Given a set of sites A and two sites v and x ,*

$$\tau_p(v, x) - \tau_p^A(v, x) = \langle I_2(v, x; A) \rangle_p + p \sum_{(y,y')} \langle I_2(v, y; A) \tau_p^{C_{\{v,y'\}}(v)}(y', x) \rangle_p. \quad (5.5.48)$$

Proof. The left side is the probability of the event that v and x are connected but are not connected in $\mathbf{Z}^d \setminus A$. By definition, this is the probability that v is connected to x through A . If v is connected to x through A then either (i) there is no pivotal bond for the connection from v to x whose first endpoint is connected to v through A , or (ii) there is such a pivotal bond. Case (i) is exactly the event $E_2(v, x; A)$, and gives the first term on the right side of (5.5.48). In case (ii), let (y, y') denote the first pivotal bond for the connection from v to x such that y is connected to v through A . The contribution to the left side of (5.5.48) due to this case is

$$\sum_{(y,y')} \langle I[E_2(v, y; A) \text{ occurs and } (y, y') \text{ is occupied and pivotal for the connection from } v \text{ to } x] \rangle_p.$$

Then by Lemma 5.5.4, with (v, y) playing the role of $(0, u)$, the contribution due to this case gives the second term on the right side of (5.5.48). $\square$

Using Lemma 5.5.5 and (5.5.46), and replacing the summation index (u, v) by (y_1, y'_1) , we have

$$\begin{aligned} R_p^{(0)}(0, x) &= p \sum_{(y_1, y'_1)} \langle I[y_1 \in D_c(0)] \langle I_2(y'_1, x; C_{\{y_1, y'_1\}}^0(0)) \rangle^{(1)} \rangle^{(0)} \\ &\quad + p^2 \sum_{(y_1, y'_1)} \sum_{(y_2, y'_2)} \langle I[y_1 \in D_c(0)] \langle I_2(y'_1, y_2; C_{\{y_1, y'_1\}}^0(0)) \rangle \\ &\quad \times \tau_p^{C_{\{y_2, y'_2\}}^1(y'_1)}(y'_2, x) \rangle^{(1)} \rangle^{(0)}. \end{aligned}$$

Here and in the following we simplify the notation by dropping the subscript p from the angular brackets denoting expectation. In addition we have

introduced a superscript to coordinate random sets with the appropriate expectation, in nested expectations. Thus for example in the second term in the right side of the above equation, the set $C_{\{y_1, y'_1\}}^0(0)$ is random with respect to the outer expectation, but may be treated as deterministic in the evaluation of the inner expectation. Using the analogue of (5.5.43) to replace the restricted two-point function on the right side by an unrestricted two-point function plus a correction, and defining

$$\Pi_p^{(1)}(0, x) = p \sum_{(y_1, y'_1)} \langle I[y_1 \in D_c(0)] \langle I_2(y'_1, x; C_{\{y_1, y'_1\}}^0(0)) \rangle^{(1)} \rangle^{(0)}$$

and

$$\begin{aligned} R_p^{(1)}(0, x) &= p^2 \sum_{(y_1, y'_1)} \sum_{(y_2, y'_2)} \langle I[y_1 \in D_c(0)] \langle I_2(y'_1, y_2; C_{\{y_1, y'_1\}}^0(0)) \\ &\quad \times \{ \tau_p(y'_2, x) - \tau_p^{C_{\{y_2, y'_2\}}^1(y'_1)}(y'_2, x) \} \rangle^{(1)} \rangle^{(0)}, \end{aligned}$$

we now have from (5.5.47) that

$$\tau_p(0, x) = g_p(0, x) - \Pi_p^{(1)}(0, x) + [g_p - \Pi_p^{(1)}] * p\Omega D * \tau_p(x) + R_p^{(1)}(0, x), \quad (5.5.49)$$

where D denotes Ω^{-1} times the indicator function of the set Ω .

The above procedure can be iterated as many times as desired. The result is the lace expansion for percolation, which is stated in the next theorem. We refrain from giving the simple but tedious details of the proof. For the statement of the theorem we write $y_0 = 0$ and introduce for $n \geq 1$,

$$C^{n-1} = C_{\{y_n, y'_n\}}^{n-1}(y'_{n-1}),$$

$$I^n = I_2(y'_n, y_{n+1}; C^{n-1}),$$

$$\begin{aligned} \Pi_p^{(n)}(0, x) &= p^n \sum_{(y_1, y'_1)} \dots \sum_{(y_n, y'_n)} \langle I[y_1 \in D_c(0)] \langle I^1 \langle I^2 \langle I^3 \dots \langle I^{n-1} \\ &\quad \times \langle I_2(y'_n, x; C^{n-1}) \rangle^{(n)} \rangle^{(n-1)} \dots \rangle^{(3)} \rangle^{(2)} \rangle^{(1)} \rangle^{(0)}, \end{aligned}$$

$$h_p^{(n)}(0, x) = g_p(0, x) + \sum_{j=1}^n (-1)^j \Pi_p^{(j)}(0, x)$$

and

$$\begin{aligned} R_p^{(n)}(0, x) &= p^{n+1} \sum_{(y_1, y'_1)} \dots \sum_{(y_{n+1}, y'_{n+1})} \langle I[y_1 \in D_c(0)] \langle I^1 \langle I^2 \dots \langle I^n \\ &\quad \times \{ \tau_p(y'_{n+1}, x) - \tau_p^{C^n}(y'_{n+1}, x) \} \rangle^{(n)} \dots \rangle^{(3)} \rangle^{(2)} \rangle^{(1)} \rangle^{(0)}. \end{aligned}$$

Finally defining $h_p^{(0)}(0, x) = g_p(0, x)$, we have the following theorem.

Theorem 5.5.6 For $p < p_c$ and $N \geq 0$,

$$\tau_p(0, x) = h_p^{(N)}(0, x) + h_p^{(N)} * p\Omega D * \tau_p(x) + (-1)^{N+1} R_p^{(N)}(0, x). \quad (5.5.50)$$

Bounds on the expansion

For effective bounds on the expansion the main requirement is to bound $\hat{g}_p(k) - 1$ and $\hat{\Pi}_p^{(n)}(k)$, as well as their second derivatives with respect to k . Bounds on the remainder term then follow readily. The method for obtaining bounds is to first obtain upper bounds in terms of Feynman diagrams, and then bound the diagrams in terms of the triangle diagram:

$$\mathbb{T}_p = \sum_{x, y \in \mathbb{Z}^d} \tau_p(0, x) \tau_p(x, y) \tau_p(y, 0) - 1. \quad (5.5.51)$$

For second derivatives with respect to k_μ , the bubble diagram with one line weighted with x_μ^2 will also occur in upper bounds. Here we will discuss the main idea used in obtaining bounds in terms of Feynman diagrams.

To obtain a diagrammatic bound on $\hat{\Pi}_p^{(n)}(k)$ for the self-avoiding walk (or for lattice trees), an important role was played by the repulsive character of the interaction: we were able to ignore interactions between distinct subwalks in upper bounds. For percolation this step is performed using the inequality of van den Berg and Kesten (1985). This is a rather general inequality, but for our purposes the following special case will suffice.

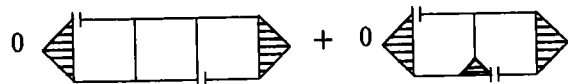
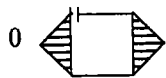
Lemma 5.5.7 Let $V_1, V_2, \dots, V_n$ be sets of paths in the lattice, and let E_i ($i = 1, \dots, n$) be the event that at least one of the paths in V_i is occupied. Let F be the event that there are pairwise disjoint occupied paths from each of the sets $V_1, V_2, \dots, V_n$. Then

$$\langle I[F] \rangle_p \leq \prod_{i=1}^n \langle I[E_i] \rangle_p.$$

It follows immediately from Lemma 5.5.7 and the definition of g_p in (5.5.45) that

$$|\hat{g}_p(k) - 1| \leq \sum_{x \neq 0} \tau_p(0, x)^2. \quad (5.5.52)$$

$$\Pi_p^{(1)}(0, x) = p \sum_{(y_1, y'_1)} \langle I[y_1 \in D_c(0)] \langle I_2(y'_1, x; C_{\{y_1, y'_1\}}^0(0)) \rangle^{(1)} \rangle^{(0)}. \quad (5.556)$$



$$\begin{aligned} \Pi_p^{(1)}(0, x) &\leq p \sum_{(y_1, y'_1)} \sum_{u, z'} \langle I[y_1 \in D_c(0)] I[u \in C_{\{y_1, y'_1\}}^0(0)] \rangle^{(0)} \\ &\quad \times \tau_p(y'_1, z') \tau_p(z', u) \tau_p(u, x) \tau_p(z', x). \end{aligned}$$

$$\begin{aligned} \Pi_p^{(1)}(0, x) &\leq p \sum_{(y_1, y'_1)} \sum_{u, w, z'} \tau_p(0, y_1) \tau_p(0, w) \tau_p(w, y_1) \tau_p(w, u) \\ &\quad \times \tau_p(y'_1, z') \tau_p(z', u) \tau_p(u, x) \tau_p(z', x). \end{aligned} \quad (5.5.57)$$

The diagrams which bound $\hat{\Pi}_p^{(n)}$ can then be estimated in terms of

$$\sup_{w \in \mathbb{Z}^d} \left[\sum_{x, y} \tau_p(0, x) \tau_p(x, y) \tau_p(y, w) - \delta_{0, w} \right],$$

which is closely related to the triangle diagram with the trivial unit contribution omitted. For further details the reader is referred to Hara and Slade (1990a).

5.6 Notes

Sections 5.1 and 5.2. The idea of expanding the self-avoidance interaction is a natural one which has occurred from time to time in the literature. However the lace expansion is the first such expansion which has been controlled rigorously. The lace expansion was first introduced in Brydges and Spencer (1985), who applied it to the weakly self-avoiding walk [defined below (5.2.2)], with interaction strength λ sufficiently small and dimension $d > 4$. In this context, they proved that the mean-square displacement is asymptotically linear in the number of steps and the scaling limit of the endpoint is Gaussian.

The connection between the lace expansion and the cluster expansions of statistical mechanics and constructive quantum field theory is discussed in Brydges (1986). The derivation of the lace expansion via the inclusion-exclusion relation, described in Section 5.1, was first given in Slade (1991). Lemma 5.2.6 was first proved in Lawler (1989).

Section 5.3. The memory-two walk has been studied by many authors. The expression (5.3.11) for the mean-square displacement was first obtained in Domb and Fisher (1958), in a more general setting. See also Barber and Ninham (1970), and Ernst (1988). For finite memory greater than two it is unclear how to evaluate $\hat{\Pi}_z(k; \tau)$ explicitly.

Section 5.4. The bounds given on the lace expansion generally follow the method of Brydges and Spencer (1985). Hara and Slade (1992b) use considerably more elaborate bounds to obtain convergence of the expansion for the strictly self-avoiding walk in five or more dimensions.

Section 5.5.1. For lattice trees and lattice animals subadditivity arguments showing that $t_n \leq \lambda^n$ and $a_n \leq \lambda_n^a$ are given in Klein (1981) and Klarner (1967) respectively. A lower bound $t_n \geq \text{const.} \lambda^n n^{-\text{const.} \log n}$ is given in Janse van Rensburg (1992b). A field theory representation was given in Lubensky and Isaacson (1979), where arguments were put forth that the upper critical dimension is eight. Skeleton inequalities were used in Bovier, Fröhlich and Glaes (1986) and Tasaki and Hara (1987) to illustrate the importance of the square diagram and the relevance of its finiteness at the critical point for mean-field behaviour.

Lace expansion methods were first applied to trees and animals in Hara and Slade (1990b). They considered the nearest-neighbour model in sufficiently high dimensions, and spread-out models above eight dimensions, and proved that the susceptibility (defined by summing the two-point function over the lattice) is bounded above and below by constant multiples of $(z_c - z)^{-1/2}$, and the correlation length of order two is bounded above and below by constant multiples of $(z_c - z)^{-1/4}$. These results were improved in Hara and Slade (1992c) to give the control of fixed- n quantities for trees stated in Theorem 5.5.1. The proof uses the fractional derivative methods of Section 6.3.

Section 5.5.2. Introductions to percolation, with varying perspectives, are given in the books by Grimmett (1989), Durrett (1988), Stauffer (1985) and Kesten (1982); see also the review article Kesten (1987). The lace expansion for percolation was introduced in Hara and Slade (1990a), where it was proven that the triangle condition is satisfied for the nearest-neighbour model above 48 dimensions (now improved to 42 — still not optimal!) and for spread-out models (of greater generality than stated in Theorem 5.5.2) above six dimensions. The triangle condition states that the triangle diagram $\sum_{x,y} \tau_p(0,x) \tau_p(x,y) \tau_p(y,0)$ is finite at the critical point $p = p_c$. It is not obvious that the triangle condition holds, since $\sum_x \tau_p(0,x)$ diverges as $p \nearrow p_c$. The triangle condition was first introduced by Aizenman and Newman (1984) as a sufficient condition for mean-field behaviour of the susceptibility (expected cluster size). Later it was shown in Nguyen (1987) that the gap exponents also take their mean-field values if the triangle condition is satisfied. Then Barsky and Aizenman (1991) showed that if the triangle condition is satisfied then the critical exponents for the percolation probability $\theta(p)$ and the magnetization (the critical exponent δ) exist and take their mean-field values. Thus the portion of Theorem 5.5.2 corresponding to the susceptibility and the percolation probability is a combination of results of Aizenman and Newman (1984), Barsky and Aizenman (1991) and Hara and Slade (1990a). The control of the correlation length stated in Theorem 5.5.2 was later obtained in Hara (1990), using a modification of the lace expansion method. Section 6.5.1 describes an analogous argument for the self-avoiding walk correlation length. A further result for the critical behaviour of high dimensional percolation was obtained in Yang and Zhang (1992), who showed that given a nonnegative integer n the cluster density function $\kappa(p)$ has a finite n -th left derivative at p_c , if d is sufficiently large.

Combining the result of Hara (1990) with the upper and lower bounds in terms of the percolation correlation length for the time constant for first passage percolation [due respectively to Chayes (1991) and Chayes and Chayes (1986a)], the time constant for the nearest-neighbour model in high

dimensions is seen to behave like $(p_c - p)^{1/2}$.

It follows from Theorem 5.5.2 that under the hypotheses of the theorem the percolation probability is continuous at the critical point in the sense that $\theta(p_c) = 0$. Although this is very strongly believed to be true in all dimensions, the only other instance of a rigorous proof that the critical percolation probability vanishes is for the nearest-neighbour model in two dimensions [see Grimmett (1989) for a proof and references to original work of Harris and Kesten]. Continuity at all points other than p_c is proven in Aizenman, Kesten and Newman (1987).

The lace expansion also leads to a bound on the nearest-neighbour percolation critical point of the form $(2d - 1)^{-1} \leq p_c \leq (2d)^{-1} + O(d^{-2})$; see Hara and Slade (1990a). This improves an estimate of Kesten (1990). For the spread-out model in more than six dimensions it is implicit in Hara and Slade (1990a) that the critical point satisfies $\Omega^{-1} \leq p_c \leq \Omega^{-1} + o(L^{-d-2})$. For spread-out percolation in general dimensions $d \geq 2$, Penrose (1992) has shown that $p_c \sim \Omega^{-1}$ as $L \rightarrow \infty$.

As stated, Lemma 2.1 of Hara and Slade (1990a) is incorrect: the class of events in the statement is too large. However the conclusion of the lemma is true for the events to which the lemma is applied. The error, which was pointed out by Y. Higuchi, is corrected in Lemma 5.5.4. Lemma 5.5.4 considers a slightly different event than that used in Hara and Slade (1990a), to simplify the presentation.

The lace expansion has also been applied to oriented percolation models, in which bonds are oriented in one direction. The upper critical dimension for these models is believed to be five [Obukhov (1980)]. Strong evidence in favour of this belief is given by the results of Nguyen and Yang (1991), who proved that the triangle condition holds for the nearest-neighbour model in very high dimensions or for spread-out models above five dimensions. This implies mean-field behaviour for various critical exponents as for the unoriented case. It appears that the lace expansion can also be applied to prove Gaussian behaviour of the two-point function at the critical point, at least for the nearest-neighbour model in high dimensions [Nguyen and Yang (in preparation)].

self-avoidance constraint would merely discourage such steps. This leads us to consider a nearest-neighbour walk, starting at the origin, with transition probabilities

$$P(\omega(n+1) = x+a | \omega(n) = x) = \frac{e^{-\lambda N_{x+a}}}{\sum_{b:|b|=1} e^{-\lambda N_{x+b}}} \quad (10.4.1)$$

where $|a| = 1$, $\lambda \geq 0$ represents the strength of the repulsion, and N_u denotes the number of visits to the site u up to time n . The case $\lambda = \infty$ then corresponds to the prohibitive model introduced in the previous paragraph.

There are as yet no rigorous results concerning the critical behaviour of the myopic self-avoiding walk. However the nonrigorous results indicate that the myopic self-avoiding walk behaves quite differently from the self-avoiding walk. Both field theoretic methods [Amit, Parisi and Peliti (1983)] and calculations related to those leading to the Flory exponents for the self-avoiding walk [Pietronero (1983)] point to an upper critical dimension of two. The diffusive behaviour $\nu = 1/2$ is expected above two dimensions, logarithmic corrections to diffusive behaviour are obtained in two dimensions, and the exponent $\nu = 2/3$ is found in one dimension. The claim that $\nu = 2/3$ in one dimension clearly does not apply when $\lambda = \infty$, for which the myopic walk behaves ballistically and $\nu = 1$. This indicates that the $\lambda = \infty$ walk belongs to a different universality class than the finite λ version; however the nonrigorous results appear to claim that the upper critical dimension is two for all $\lambda \leq \infty$. A survey, with references to numerical calculations, is given in Peliti and Pietronero (1987).

Appendix A

Random walk

This appendix contains a number of elementary facts about the usual random walk with no self-avoidance constraint, which are used in the book. Further details on random walk can be found in Lawler (1991), Spitzer (1976), and Montroll and West (1979).

We consider a random walk taking steps in a set Ω which does not contain the origin and which is invariant under the symmetries of the lattice $\mathbb{Z}^d$. The two basic examples we have in mind here are Ω equal to the set of nearest neighbours of the origin, and $\Omega = \{x \in \mathbb{Z}^d : 0 < \|x\|_\infty \leq L\}$. As usual we denote the cardinality of Ω also by Ω . The two-point function is defined by

$$C_z(x, y) = \sum_{\omega: x \rightarrow y} z^{|\omega|}, \quad (A.1)$$

where z is a complex parameter and the sum is over random walks of arbitrary length, taking steps in Ω , which begin at x and end at y . Clearly $C_z(x, y) = C_z(0, y-x)$. If $x = y$ then the two-point function includes a unit contribution from the zero-step walk. Since the total number of n -step random walks beginning at x is equal to Ω^n , $\sum_x |C_z(0, x)| \leq \sum_{n=0}^{\infty} (\Omega|z|)^n$ and hence the two-point function and its Fourier transform $\hat{C}_z(k)$ [defined in (1.4.10)] are both finite for $|z| < \Omega^{-1}$. For $z = \Omega^{-1}$, (A.1) is the Green function

$$C_{1/\Omega}(x, y) = \sum_{n=0}^{\infty} p_n(x, y), \quad (A.2)$$

where $p_n(x, y)$ is the probability that the walk starting at x is at y after n steps. We will show momentarily that (A.2) is finite for $z = \Omega^{-1}$ if $d > 2$.

Fix z with $|z| < \Omega^{-1}$. Extracting from the two-point function the contribution due to the zero-step walk, and taking into account all possible

locations of the walk after the first step, we obtain

$$C_z(0, x) = \delta_{0, x} + \sum_{y \in \Omega} z \sum_{\omega: y \rightarrow x} z^{|\omega|}. \quad (\text{A.3})$$

This can be rewritten as the convolution equation

$$C_z(0, x) = \delta_{0, x} + z\Omega D * C_z(x), \quad (\text{A.4})$$

where

$$D(x) = \begin{cases} \Omega^{-1}, & x \in \Omega \\ 0, & x \notin \Omega. \end{cases} \quad (\text{A.5})$$

Taking the Fourier transform of (A.4) gives

$$\hat{C}_z(k) = \frac{1}{1 - z\Omega \hat{D}(k)}, \quad k \in [-\pi, \pi]^d \quad (\text{A.6})$$

with

$$\hat{D}(k) = \frac{1}{\Omega} \sum_{y \in \Omega} e^{ik \cdot y} = \frac{1}{\Omega} \sum_{y \in \Omega} \cos k \cdot y. \quad (\text{A.7})$$

Taking the inverse transform then gives

$$C_z(0, x) = \int_{[-\pi, \pi]^d} \frac{e^{-ik \cdot x}}{1 - z\Omega \hat{D}(k)} \frac{d^d k}{(2\pi)^d}. \quad (\text{A.8})$$

Since $1 - \hat{D}(k)$ behaves like k^2 near $k = 0$, it follows from (A.8) and a limiting argument that $C_{1/\Omega}(x, y)$ is finite if $d > 2$.

The critical exponents for random walk are much easier to determine than those for the self-avoiding walk. It follows immediately from the facts that there are Ω^n n -step random walks starting from the origin and that the mean-square displacement after n steps is proportional to n that $\gamma = 1$ and $\nu = 1/2$. Also, $\alpha_{sing} - 2 = -d/2$ follows from the local central limit theorem, which states that the probability of ending at x after n steps (with n and $\|x\|_1$ of the same parity, x fixed, and $n \rightarrow \infty$) is asymptotically of order $n^{-d/2}$. The value $\eta = 0$ can be seen in k -space from the fact that $\hat{C}_{1/\Omega}(k)^{-1} = \text{const.} k^2 + O(k^4)$. In x -space the analogous statement follows from the well-known $|x|^{2-d}$ behaviour of the Green function $C_{1/\Omega}(0, x)$. Unfortunately none of these provides a useful means of understanding the corresponding critical exponents for the self-avoiding walk.

However, for the correlation length the proof of mean-field behaviour for the self-avoiding walk in high dimensions given in Section 6.5.1 is modelled on a corresponding argument for random walk (with considerable additional

input). Because this argument is not a standard one, we give it below to prove that the nearest-neighbour (or *simple*) random walk correlation length has critical exponent $1/2$. This may serve to motivate the approach used in Section 6.5.1. One ingredient is the next lemma, which is of independent interest. The lemma states quite generally that if a two-point function satisfies a certain inequality, known as the *Lieb-Simon inequality* [Simon (1980), Lieb (1980)], then it decays exponentially if the susceptibility is finite.

In the statement of the lemma, B_R is the set of all $y \in \mathbb{Z}^d$ such that $\|y\|_\infty \leq R$, and ∂B_R is defined to be the set of sites y with $\|y\|_\infty = R$.

Lemma A.1 Suppose $K(x, y)$ is nonnegative and translation invariant, and satisfies

$$K(0, x) \leq \sum_{y \in \partial B_R} K(0, y) K(y, x), \quad (\text{A.9})$$

for all $R \geq 1$ and $x \notin B_R$. If $\sum_x K(0, x) < \infty$, then there are constants $C, m > 0$ such that $K(0, x) \leq C e^{-m|x|}$ for all sites x .

Proof. Fix any θ with $0 < \theta < 1$. Since $\sum_x K(0, x) < \infty$, there is an R_0 such that for $R \geq R_0$

$$\sum_{y \in \partial B_R} K(0, y) \leq \theta < 1.$$

Fix an x with $\|x\|_\infty > R_0$, and let $n = \lfloor \|x\|_\infty / R_0 \rfloor$. Then $n \geq 1$. It follows from (A.9) that

$$K(0, x) \leq \theta K(0, x - y_1), \quad (\text{A.10})$$

where y_1 is a site in ∂B_R which satisfies $K(y_1, x) = \sup_{y \in \partial B_{R_0}} K(y, x)$. Clearly $\|x - y_1\|_\infty \geq \|x\|_\infty - R_0 \geq (n - 1)R_0$. Now we iterate (A.10) $n - 2$ more times to obtain

$$K(0, x) \leq \theta^{n-1} K(0, x - y_{n-1}),$$

for some site y_n . Since $n \geq \frac{\|x\|_\infty}{R_0} - 1$, this gives

$$K(0, x) \leq \left(\theta^{\frac{1}{R_0}}\right)^{\|x\|_\infty} \theta^{-2} \sup_y K(0, y).$$

The desired exponential decay now follows from the fact that the supremum on the right side is finite by hypothesis, together with the inequality $\|x\|_\infty \geq d^{-1/2}|x|$. $\square$

We now apply this lemma to show that the random walk correlation length has critical exponent $1/2$. For simplicity we consider only the simple random walk, and write $e_1 = (1, 0, \dots, 0)$.

Theorem A.2 Let Ω consist of the $2d$ nearest neighbours of the origin in $\mathbb{Z}^d$.

(a) For $0 < z < (2d)^{-1}$ the mass

$$m_0(z) \equiv - \lim_{n \rightarrow \infty} n^{-1} \log C_z(0, ne_1)$$

exists, is strictly positive and finite, and for all x

$$C_z(0, x) \leq C_z(0, 0) e^{-m_0(z) \|x\|_\infty}. \quad (\text{A.11})$$

(b) For $0 < z < (2d)^{-1}$, the simple random walk susceptibility $\chi_0(z) = \sum_x C_z(0, x)$ satisfies

$$\chi_0(z)^{-1} = 2z [\cosh m_0(z) - 1], \quad (\text{A.12})$$

and hence $m_0(z)^2 \sim z^{-1}(1 - 2dz)$ as $z \nearrow (2d)^{-1}$, i.e.,

$$\xi_0(z) = m_0(z)^{-1} \sim \left(\frac{1}{2d}\right)^{1/2} \left(\frac{1}{1 - 2dz}\right)^{1/2}.$$

Proof. (a) Assuming the limit exists, the proof from Section 1.3 that $m(z)$ is finite and nonzero applies also to $m_0(z)$ [see (1.3.14) and (1.3.16)]. We now obtain a subadditivity inequality which will yield existence of the limit. Given three sites w, x, y , any simple random walk ω from w to y which passes through x can be decomposed into two subwalks as follows. Let j be the last time that $\omega(j) = x$, and let $\omega^{(1)}$ be the initial j -step portion of ω and $\omega^{(2)}$ be the remainder. Then $\omega^{(2)}$ does not return to x . The generating function for walks from x to y which do not return to x is equal to $C_z(x, x)^{-1} C_z(x, y)$. Therefore

$$\sum_{\omega: w \rightarrow y} z^{|\omega|} I[x \in \omega] = C_z(w, x) C_z(x, x)^{-1} C_z(x, y). \quad (\text{A.13})$$

Using (A.13) and translation invariance,

$$\begin{aligned} C_z(0, (n+m)e_1) &\geq \sum_{\omega: 0 \rightarrow (n+m)e_1} z^{|\omega|} I[ne_1 \in \omega] \\ &= C_z(0, ne_1) C_z(0, 0)^{-1} C_z(0, me_1). \end{aligned} \quad (\text{A.14})$$

This implies that the sequence $-\log[C_z(0, ne_1)/C_z(0, 0)]$ is subadditive, so by Lemma 1.2.2 the limit defining m_0 exists and

$$C_z(0, ne_1) \leq C_z(0, 0) e^{-m_0(z)n}. \quad (\text{A.15})$$

To prove (A.11) we first interchange two components of x if necessary, and then change the sign of x_1 if necessary, to ensure that $x_1 = \|x\|_\infty$. This does not change the value of $C_z(0, x)$. Then symmetry and (A.13) with $w = 0$ and $y = 2\|x\|_\infty e_1$ give

$$C_z(0, 2\|x\|_\infty e_1) \geq \sum_{\omega: 0 \rightarrow 2\|x\|_\infty e_1} z^{|\omega|} I[x \in \omega] = C_z(0, x)^2 C_z(0, 0)^{-1}.$$

The bound (A.11) then follows from (A.15).

(b) To prove the identity (A.12) we introduce the exponentially weighted susceptibility

$$\chi_t(z) = \sum_x e^{tx_1} C_z(0, x). \quad (\text{A.16})$$

Inserting (A.3) into (A.16) and using translation invariance gives

$$\begin{aligned} \chi_t(z) &= 1 + \sum_{y: |y|=1} z e^{ty_1} \sum_x e^{t(x_1 - y_1)} C_z(y, x) \\ &= 1 + 2z [d - 1 + \cosh t] \chi_t(z). \end{aligned} \quad (\text{A.17})$$

Therefore

$$\chi_t(z) = \frac{1}{1 - 2z[d - 1 + \cosh t]}. \quad (\text{A.18})$$

By (A.18), $\chi_t(z)$ is an increasing function of t . By (A.11), $\chi_t(z)$ is finite if $z < 1/2d$ and $t < m_0(z)$. We now argue that for fixed $z < 1/2d$,

$$\lim_{t \nearrow m_0(z)} \chi_t(z) = \infty.$$

To see this, suppose to the contrary that $\chi_t(z)$ is uniformly bounded for $t < m_0(z)$. Then by the monotone convergence theorem $\chi_{m_0(z)}(z) < \infty$. Let $K(0, x) = e^{m_0(z)x_1} C_z(0, x)$. This K satisfies the Lieb-Simon inequality (A.9), as can be seen using the fact that a walk from 0 to $x \notin B_R$ must hit a site in ∂B_R . Hence by Lemma A.1 $e^{m_0(z)x_1} C_z(0, x)$ decays exponentially, which contradicts the definition of $m_0(z)$.

By (A.18),

$$\chi_0(z)^{-1} - \chi_t(z)^{-1} = 2z [\cosh t - 1].$$

Taking the limit $t \nearrow m_0(z)$ gives

$$\chi_0(z)^{-1} = 2z [\cosh m_0(z) - 1].$$

This gives the desired result

$$m_0(z)^2 \sim \frac{1 - 2dz}{z} \quad \text{as } z \nearrow \frac{1}{2d},$$

and completes the proof. $\square$

We now turn to two lemmas which are needed in Section 6.2 to prove convergence of the lace expansion. These lemmas give estimates for the two-point function for two particular choices of the step set Ω . We begin with the nearest-neighbour walk, for which the set Ω consists of the $2d$ nearest neighbours of the origin. In this case

$$\hat{D}(k) = \hat{D}_0(k) \equiv \frac{1}{d} \sum_{\mu=1}^d \cos k_{\mu}. \quad (\text{A.19})$$

We will denote the two-point function for the nearest-neighbour model by $C_z^{(0)}(0, x)$. The next lemma gives estimates for $\hat{D}_0(k)$, and for $\hat{C}_z^{(0)}(k)$ at the critical point $z = \Omega^{-1}$. Although for our purposes the dimension d is always a positive integer, bounds have been given in terms of $(d-4)^{-1}$ rather than d^{-1} to emphasize critical dimensionality. For a different approach to some bounds of this type, see Hara (1990).

Lemma A.3 For all $d \geq 1$ and $k \in [-\pi, \pi]^d$,

$$1 - \hat{D}_0(k) \geq \frac{2}{\pi^2 d} k^2. \quad (\text{A.20})$$

For all nonnegative integers n and all $d \geq 1$,

$$\|\hat{D}_0^n\|_1 \leq \left(\frac{\pi d}{4n}\right)^{d/2}. \quad (\text{A.21})$$

There is a constant K , independent of d , such that for $z = (2d)^{-1}$ and for all $d > 4$

$$0 \leq \|\hat{C}^{(0)}\|_2^2 - 1 \leq K(d-4)^{-1} \quad (\text{A.22})$$

and

$$\left\| \frac{\partial_{\mu}^2 \hat{D}_0}{[1 - \hat{D}_0]^2} \right\|_1 + 2 \left\| \frac{(\partial_{\mu} \hat{D}_0)^2}{[1 - \hat{D}_0]^3} \right\|_1 \leq K(d-4)^{-1}. \quad (\text{A.23})$$

The above norms are L^p norms on $[-\pi, \pi]^d$ with measure $(2\pi)^{-d} d^d k$.

Proof. We consider each inequality in turn. The first follows from the fact that $1 - \cos k_{\mu} \geq 2\pi^{-2} k_{\mu}^2$, for $|k_{\mu}| \leq \pi$.

For the second inequality, we proceed as follows. By symmetry,

$$\|\hat{D}_0^n\|_1 = \pi^{-d} \int_{[0, \pi]^d} |\hat{D}_0(k)|^n d^d k \quad (\text{A.24})$$

$$= \pi^{-d} \sum_{m=0}^d \binom{d}{m} \int_{[0, \pi/2]^m} dk_1 \dots dk_m \times \int_{[\pi/2, \pi]^{d-m}} dk_{m+1} \dots dk_d |\hat{D}_0(k)|^n.$$

For k in the domain of integration on the right side,

$$\begin{aligned} |\hat{D}_0(k)| &\leq d^{-1} \sum_{\mu=1}^d |\cos k_{\mu}| \\ &\leq 1 - \frac{4}{\pi^2 d} \left[\sum_{\mu=1}^m k_{\mu}^2 + \sum_{\nu=m+1}^d (\pi - k_{\nu})^2 \right] \\ &\leq \exp \left(-\frac{4}{\pi^2 d} \left[\sum_{\mu=1}^m k_{\mu}^2 + \sum_{\nu=m+1}^d (\pi - k_{\nu})^2 \right] \right). \end{aligned} \quad (\text{A.25})$$

Combining (A.24) and (A.25) and making a change of variables gives

$$\|\hat{D}_0^n\|_1 \leq \left(\frac{2}{\pi}\right)^d \int_{[0, \pi/2]^d} \exp[-4nk^2 \pi^{-2} d^{-1}] d^d k. \quad (\text{A.26})$$

Extending the domain of integration, we obtain

$$\|\hat{D}_0^n\|_1 \leq \pi^{-d} \int_{\mathbb{R}^d} \exp[-4nk^2 \pi^{-2} d^{-1}] d^d k = \left(\frac{\pi d}{4n}\right)^{d/2}. \quad (\text{A.27})$$

For the bound on $\|\hat{C}^{(0)}\|_2^2 - 1$, we use a series expansion of $\hat{C}^{(0)}(k)^2 = [1 - \hat{D}_0(k)]^{-2}$ [which follows from (A.6)] to obtain

$$\|\hat{C}^{(0)}\|_2^2 = 1 + \sum_{n=1}^{\infty} (n+1) \hat{D}_0^n \frac{d^d k}{(2\pi)^d}. \quad (\text{A.28})$$

The integral here is over $[-\pi, \pi]^d$. It suffices to show that there is a constant K (independent of d) such that for $d > 4$,

$$\sum_{n=1}^{\infty} (n+1) \int \hat{D}_0^n \frac{d^d k}{(2\pi)^d} \leq K(d-4)^{-1}. \quad (\text{A.29})$$

For n odd the integral is zero, while for each even n the integral is positive.

To prove (A.29), we divide the summation into two parts and use (A.27) to obtain

$$\sum_{n=1}^{\infty} (n+1) \int \hat{D}_0^n \frac{d^d k}{(2\pi)^d} \leq \sum_{n=1}^{d-1} (n+1) \int \hat{D}_0^n \frac{d^d k}{(2\pi)^d} + \sum_{n=d}^{\infty} (n+1) \left(\frac{\pi d}{4n}\right)^{d/2}. \quad (\text{A.30})$$

Using the fact that for $a > 1$

$$\sum_{n=N}^{\infty} \frac{1}{n^a} \leq \int_{N-1}^{\infty} \frac{1}{t^a} dt = \frac{(N-1)^{-a+1}}{a-1},$$

it is not hard to see that the second sum on the right side of (A.30) is bounded by a dimension-independent multiple of $(d-4)^{-1}$. For the first sum on the right side, we use the fact that $|\hat{D}_0(k)| \leq 1$ to obtain

$$\sum_{n=1}^{d-1} (n+1) \int \hat{D}_0^n \frac{d^d k}{(2\pi)^d} \leq 3 \int \hat{D}_0^2 \frac{d^d k}{(2\pi)^d} + 5 \int \hat{D}_0^4 \frac{d^d k}{(2\pi)^d} + d^2 \int \hat{D}_0^6 \frac{d^d k}{(2\pi)^d}. \quad (\text{A.31})$$

The right side can be estimated by observing that the integrals on the right side are respectively the probabilities that simple random walk returns to the origin after two, four, or six steps. These probabilities are respectively of the order of d^{-1} , d^{-2} and d^{-3} , which yields (A.29).

For (A.23) we proceed as follows. We first observe that integration by parts gives

$$\left\| \frac{(\partial_\mu \hat{D}_0)^2}{[1 - \hat{D}_0]^3} \right\|_1 = -\frac{1}{2} \int \frac{\partial_\mu^2 \hat{D}_0}{[1 - \hat{D}_0]^2} \frac{d^d k}{(2\pi)^d}, \quad (\text{A.32})$$

so it suffices to bound the first term on the left side of (A.23). But for this we simply observe that

$$\partial_\mu^2 \hat{D}_0(k) = -d^{-1} \cos k_\mu, \quad (\text{A.33})$$

which is bounded in absolute value by d^{-1} , and then use (A.22). $\square$

The second model of ordinary random walk we consider is the “spread-out” walk in $\mathbf{Z}^d$ with step set $\Omega = \{x \neq 0 : \|x\|_\infty \leq L\}$. The cardinality of this set is $\Omega = (2L+1)^d - 1$. In this case we will be interested in $d > 4$ and large L . We write $\hat{D}_L(k)$ in place of $\hat{D}(k)$ of (A.7) for this model. Also, we write $C^{(L)}(0, x)$ for the critical ($z = \Omega^{-1}$) spread-out random walk two-point function. Before proving an analogue of Lemma A.3 for this spread-out random walk, we need the following result.

Lemma A.4 For any $S \subset \{1, \dots, d\}$, $L \geq 1$ and $k \in [-\pi, \pi]^d$,

$$|\hat{D}_L(k)| \leq 2 \frac{\Omega+1}{\Omega} \left| \prod_{\mu \in S} (2L+1) \sin(k_\mu/2) \right|^{-1} \quad (\text{A.34})$$

and

$$|\partial_\mu^2 \hat{D}_L(k)| \leq \frac{\Omega+1}{\Omega} (2L+1)^2 \left| \prod_{\nu \in S} (2L+1) \sin(k_\nu/2) \right|^{-1}. \quad (\text{A.35})$$

Proof. In terms of the Dirichlet kernel

$$M(t) = \sum_{j=-L}^L e^{ijt} = \frac{\sin[(2L+1)t/2]}{\sin(t/2)}, \quad (\text{A.36})$$

we have

$$\hat{D}_L(k) = \frac{1}{\Omega} \left[(2L+1)^d \prod_{\mu=1}^d \frac{M(k_\mu)}{2L+1} - 1 \right]. \quad (\text{A.37})$$

Since $|M(t)|$ is bounded above by both $2L+1$ and $|\sin(t/2)|^{-1}$,

$$\left| \prod_{\mu=1}^d \frac{M(k_\mu)}{2L+1} \right| \leq \left| \prod_{\mu \in S} \frac{1}{(2L+1) \sin(k_\mu/2)} \right|. \quad (\text{A.38})$$

This gives (A.34), once we observe that

$$1 \leq \frac{(2L+1)^d}{\prod_{\mu \in S} |(2L+1) \sin(k_\mu/2)|}. \quad (\text{A.39})$$

For (A.35), we have

$$\partial_\mu^2 \hat{D}_L(k) = \frac{\Omega+1}{\Omega} \frac{\partial_\mu^2 M(k_\mu)}{2L+1} \prod_{\nu \neq \mu} \frac{M(k_\nu)}{2L+1}. \quad (\text{A.40})$$

In view of the above bounds, it suffices to show that $\partial_\mu^2 M(k_\mu)$ is bounded above by both $(2L+1)^3$ and $(2L+1)^2 |\sin(k_\mu/2)|^{-1}$. The first of these bounds follows from

$$\partial_\mu^2 M(k_\mu) = - \sum_{x_\mu=-L}^L x_\mu^2 e^{ik_\mu x_\mu} \quad (\text{A.41})$$

by taking absolute values inside the sum and using $x_\mu^2 \leq L^2$. For the second bound, we use the summation by parts formula

$$\sum_{n=A}^B (b_{n+1} - b_n) a_n = - \sum_{n=A}^B b_n (a_n - a_{n-1}) + b_{B+1} a_B - b_A a_{A-1}, \quad (\text{A.42})$$

to obtain

$$\begin{aligned} \sum_{x_\mu=-L}^L x_\mu^2 e^{ik_\mu x_\mu} &= \frac{e^{-ik_\mu/2}}{2i \sin(k_\mu/2)} \sum_{x_\mu=-L}^L [e^{ik_\mu(x_\mu+1)} - e^{ik_\mu x_\mu}] x_\mu^2 \\ &= -\frac{e^{-ik_\mu/2}}{2i \sin(k_\mu/2)} \left\{ \sum_{x_\mu=-L}^L e^{ik_\mu x_\mu} [x_\mu^2 - (x_\mu - 1)^2] \right. \\ &\quad \left. + e^{ik_\mu(L+1)} L^2 - e^{-ik_\mu L} (-L - 1)^2 \right\}. \end{aligned} \quad (\text{A.43})$$

The desired bound then follows. $\square$

The next lemma states several properties of the two-point function for the spread-out walk.

Lemma A.5 *For any $d \geq 1$ there is an L_0 depending only on d such that for any $k \in [-\pi, \pi]^d$ and $L \geq L_0$,*

$$1 - \hat{D}_L(k) \geq \frac{k^2}{2\pi^2 d}. \quad (\text{A.44})$$

For any $d \geq 1$ there are L_0 and K depending only on d such that for all integers $n \geq 2$ and all $L \geq L_0$,

$$\|\hat{D}_L^n\|_1 \leq K \frac{(\log L)^{d/2}}{L^d n^{d/2}}. \quad (\text{A.45})$$

Fix $z = \Omega^{-1}$. For any $d > 4$ and for all L , there is a K depending only on d such that

$$0 \leq \|\hat{C}^{(L)}\|_2^2 - 1 \leq K \frac{(\log L)^{d/2}}{L^d} \quad (\text{A.46})$$

and

$$\left\| \frac{\partial_\mu^2 \hat{D}_L}{[1 - \hat{D}_L]^2} \right\|_1 + 2 \left\| \frac{(\partial_\mu \hat{D}_L)^2}{[1 - \hat{D}_L]^3} \right\|_1 \leq K \frac{(\log L)^d}{L^{d-2}}. \quad (\text{A.47})$$

The above norms are L^p norms on $[-\pi, \pi]^d$ with measure $(2\pi)^{-d} d^d k$.

Proof. We shall begin by proving

$$1 - |\hat{D}(k)| \geq \frac{k^2}{2\pi^2 d}, \quad (\text{A.48})$$

which in particular implies (A.44). By Lemma A.4 (with $|S| = 1$), there is a universal constant a such that

$$|\hat{D}_L(k)| \leq \frac{a}{\|k\|_\infty L}. \quad (\text{A.49})$$

For $\|k\|_\infty \geq 2a/L$, this together with the fact that $k^2 \leq d\pi^2$ gives

$$1 - |\hat{D}_L(k)| \geq 1 - \frac{1}{2} \geq \frac{k^2}{2\pi^2 d}. \quad (\text{A.50})$$

For $\|k\|_\infty \leq 2a/L$, we will consider separately the two cases $\hat{D}_L(k) \geq 0$ and $\hat{D}_L(k) < 0$. In either case, it follows from symmetry and the fact that $1 - \cos t \geq 2\pi^{-2} t^2$ for $|t| \leq \pi$ that

$$\begin{aligned} 1 - \hat{D}_L(k) &\geq \frac{1}{\Omega} \sum_{x \in \Omega: |k \cdot x| \leq \pi} \frac{2}{\pi^2} (k \cdot x)^2 \\ &\geq \frac{2}{\pi^2 \Omega} \sum_{x \in \Omega: \|x\|_1 \leq \pi L/2a} (k \cdot x)^2 \\ &= \frac{2}{\pi^2 \Omega} k^2 \sum_{x \in \Omega: \|x\|_1 \leq \pi L/2a} x_1^2. \end{aligned} \quad (\text{A.51})$$

The sum over x on the right side is bounded below by a (dimension-dependent) multiple of L^{2+d} , for L sufficiently large, and hence there is a positive constant c (depending only on d) such that for $L \geq (2\pi^2 dc)^{-1/2}$ and $\|k\|_\infty \leq 2a/L$,

$$1 - \hat{D}_L(k) \geq cL^2 k^2 \geq \frac{k^2}{2\pi^2 d}. \quad (\text{A.52})$$

This gives (A.48) for $\|k\|_\infty \leq 2a/L$ and $\hat{D}_L(k) \geq 0$. If on the other hand $\hat{D}_L(k) < 0$ and $\|k\|_\infty \leq 2a/L$, then there are positive constants c, c' (depending only on the dimension) such that for large L

$$\begin{aligned} 1 - |\hat{D}_L(k)| &= \frac{1}{\Omega} \sum_{x \in \Omega} (1 + \cos k \cdot x) \\ &\geq \frac{1}{\Omega} \sum_{x \in \Omega: |k \cdot x| \leq \pi/2} 1 \\ &\geq \frac{1}{\Omega} \sum_{x \in \Omega: \|x\|_1 \leq \pi L/4a} 1 \\ &\geq c' \geq cL^2 k^2. \end{aligned} \quad (\text{A.53})$$

Taking $L \geq (2\pi^2 dc)^{-1/2}$, it follows from (A.52) and (A.53) that for $\|k\|_\infty \leq 2a/L$

$$1 - |\hat{D}_L(k)| \geq cL^2 k^2 \geq \frac{k^2}{2\pi^2 d}. \quad (\text{A.54})$$

With (A.50), this gives (A.48) and hence (A.44).

Turning now to (A.45), it follows from (A.50) and (A.54) that if L is sufficiently large then

$$\|\hat{D}_L^n\|_1 \leq \int_{\|k\|_\infty \leq 2a/L} e^{-cnL^2k^2} \frac{d^d k}{(2\pi)^d} + \frac{1}{2^n} \leq \frac{K}{2L^d n^{d/2}} + \frac{1}{2^n}, \quad (\text{A.55})$$

where K depends only on d . Fix $b \in (1, 2)$. Increasing K if necessary, for all $n \geq d \log L / \log b$ we have

$$\frac{1}{2^n} \leq \frac{K}{2b^n n^{d/2}} \leq \frac{K}{2L^d n^{d/2}}. \quad (\text{A.56})$$

Hence for $n \geq d \log L / \log b$,

$$\|\hat{D}_L^n\|_1 \leq KL^{-d} n^{-d/2}. \quad (\text{A.57})$$

On the other hand, for any $n \geq 2$, $\|\hat{D}_L^n\|_1 \leq \|\hat{D}_L^2\|_1 = \Omega^{-1}$, since $|\hat{D}_L(k)| \leq 1$ and $\|\hat{D}_L^2\|_1$ is the probability of return to the origin after two steps. Hence for all $n \leq d \log L / \log b$

$$\|\hat{D}_L^n\|_1 \leq \frac{1}{\Omega} \leq \frac{1}{\Omega} \left(\frac{d \log L}{n \log b} \right)^{d/2} \leq K(\log L)^{d/2} L^{-d} n^{-d/2}. \quad (\text{A.58})$$

This completes the proof of (A.45).

The proof of (A.46) now proceeds by writing the left side in terms of a sum of terms of order $(n+1) \int \hat{D}^n$ as in (A.28), and using (A.45).

For (A.47), just as for the nearest-neighbour walk an integration by parts allows the second term on the left to be bounded by the first. The first term is equal to

$$\sum_{n=1}^{\infty} n \int |\partial_\mu^2 \hat{D}_L(k)| \hat{D}_L(k)^{n-1} \frac{d^d k}{(2\pi)^d}. \quad (\text{A.59})$$

By (A.45) and the fact that $|\partial_\mu^2 \hat{D}_L(k)| = O(L^2)$ it suffices to bound the first two terms in the sum, and for these it is sufficient to show that

$$\int |\partial_\mu^2 \hat{D}_L(k)| \frac{d^d k}{(2\pi)^d} \leq K(\log L)^d L^{2-d}. \quad (\text{A.60})$$

To prove (A.60), we begin by dividing the domain of integration to obtain

$$\begin{aligned} \int |\partial_\mu^2 \hat{D}_L(k)| \frac{d^d k}{(2\pi)^d} &= \int_{\|k\|_\infty \leq 2a/L} |\partial_\mu^2 \hat{D}_L(k)| \frac{d^d k}{(2\pi)^d} \\ &\quad + \int_{2a/L \leq \|k\|_\infty \leq \pi} |\partial_\mu^2 \hat{D}_L(k)| \frac{d^d k}{(2\pi)^d}. \end{aligned} \quad (\text{A.61})$$

The first term on the right side is order L^{2-d} , since the integrand is order L^2 and the volume is order L^{-d} . The domain of integration of the second term is the disjoint union over nonempty subsets $S \subset \{1, \dots, d\}$ of

$$R_S = \{k \in \mathbb{R}^d : 2a/L < k_\mu \leq \pi \text{ for } \mu \in S, |k_\nu| \leq 2a/L \text{ for } \nu \notin S\}. \quad (\text{A.62})$$

By (A.35),

$$\begin{aligned} \int_{R_S} |\partial_\mu^2 \hat{D}_L(k)| \frac{d^d k}{(2\pi)^d} &\leq KL^{2-|S|} \int_{R_S} \prod_{\mu \in S} |k_\mu|^{-1} \frac{d^d k}{(2\pi)^d} \\ &\leq KL^{2-|S|-(d-|S|)} [\log L]^{|S|}. \end{aligned} \quad (\text{A.63})$$

The desired result then follows by summing over S . $\square$

We conclude this appendix with a theorem about ordinary random walks with general step distribution that is needed in Section 8.1. As explained in that section, this theorem is true under weaker hypotheses than we state here.

Theorem A.6 *Let p be a positive integer. Let $Y_1, Y_2, \dots$ be independent, identically distributed $\mathbb{Z}^p$ -valued random variables with the property that*

$$\beta \equiv \min\{\Pr\{Y_1 - Y_2 = e\} : \|e\|_1 = 1\} > 0.$$

Then there exists a finite constant C (depending on the distribution of Y_1) such that

$$\Pr\{Y_1 + \dots + Y_m = x\} \leq Cm^{-p/2} \quad \text{for every } m \geq 1, x \in \mathbb{Z}^p. \quad (\text{A.64})$$

Proof. For $k \in [-\pi, \pi]^p$, let $\phi(k) = E[\exp(ik \cdot Y_1)]$ denote the characteristic function (Fourier transform) of Y_1 . Let $e_1, \dots, e_p$ denote the positive unit vectors of $\mathbb{Z}^p$. Then by independence and symmetry,

$$\begin{aligned} |\phi(k)|^2 &= E[\exp(ik \cdot [Y_1 - Y_2])] \\ &= E[\cos(k \cdot [Y_1 - Y_2])] \\ &\leq 2 \sum_{j=1}^p \Pr\{Y_1 - Y_2 = e_j\} \cos k_j + \Pr\{|Y_1 - Y_2| \neq 1\} \\ &= 2 \sum_{j=1}^p \Pr\{Y_1 - Y_2 = e_j\} (\cos k_j - 1) + 1 \\ &\leq 2 \sum_{j=1}^p \beta (\cos k_j - 1) + 1. \end{aligned} \quad (\text{A.65})$$

Since $\cos k_j - 1 \leq -2k_j^2/\pi^2$ for every $k_j \in [-\pi, \pi]$, we have

$$|\phi(k)|^2 \leq 1 - 4\beta k^2/\pi^2 \leq e^{-\delta|k|^2} \quad (\text{A.66})$$

for every $k \in [-\pi, \pi]^p$, where $\delta = 4\beta/\pi^2 > 0$.

Next let $n = \lfloor m/2 \rfloor$, so that m is either $2n$ or $2n+1$. Let $x \in \mathbb{Z}^p$. Then by Fourier inversion and (A.66) we obtain

$$\begin{aligned} \Pr\{Y_1 + \cdots + Y_m = x\} &= \int_{[-\pi, \pi]^p} e^{-ik \cdot x} \phi(k)^m \frac{d^p k}{(2\pi)^p} \\ &\leq \int_{[-\pi, \pi]^p} |\phi(k)|^{2n} \frac{d^p k}{(2\pi)^p} \\ &\leq \int_{\mathbb{R}^p} e^{-n\delta k^2} \frac{d^p k}{(2\pi)^p} \\ &= \text{const. } n^{-p/2}. \end{aligned}$$

Since $n = \lfloor m/2 \rfloor$, this proves the theorem. $\square$

Appendix B

Proof of the renewal theorem

In this appendix we prove the version of the renewal theorem that is stated in Theorem 4.2.2, which is all that we need in this book. The probabilistic interpretation of this theorem is discussed briefly just before Theorem 4.2.2. The proof is essentially the standard one [as in for example Feller (1968) or Karlin and Taylor (1975)]. A very different proof using complex analysis appears in Kingman (1972), pp. 10-14.

Theorem B.1 Assume that $\{f_n : n \geq 1\}$ and $\{g_n : n \geq 0\}$ are nonnegative sequences, and let

$$f = \sum_{n=1}^{\infty} f_n \quad \text{and} \quad g = \sum_{n=0}^{\infty} g_n$$

denote their sums. Assume that $0 < g < +\infty$. Also assume that $f_1 > 0$. Define the new sequence $v_0, v_1, \dots$ by

$$\begin{aligned} v_0 &= g_0 \\ v_n &= g_n + f_1 v_{n-1} + f_2 v_{n-2} + \cdots + f_n v_0, \text{ for all } n \geq 1. \end{aligned} \quad (\text{B.1})$$

- (a) If $f < 1$, then $\lim_{n \rightarrow \infty} v_n = 0$ and $\sum_{n=0}^{\infty} v_n = g/(1-f)$.
 (b) If $f = 1$, then

$$\lim_{n \rightarrow \infty} v_n = \frac{g}{\sum_{k=1}^{\infty} k f_k} \quad (\text{B.2})$$

(the limit is 0 if the sum in the denominator diverges). Also, $\sum_n v_n$ diverges.

- (c) If $1 < f \leq +\infty$, then $\limsup_{n \rightarrow \infty} v_n^{1/n} > 1$.

Bibliography

- [1] J. Adler (1984). A second look at a controversial percolation exponent — is η negative in three dimensions? *Z. Phys. B - Cond. Matt.*, **55**:227–229.
- [2] R. Ahlberg and S. Janson (1980). Upper bounds for the connective constant. Unpublished manuscript.
- [3] M. Aizenman (1982). Geometric analysis of φ^4 fields and Ising models, Parts I and II. *Commun. Math. Phys.*, **86**:1–48.
- [4] M. Aizenman (1985). The intersection of Brownian paths as a case study of a renormalization group method for quantum field theory. *Commun. Math. Phys.*, **97**:91–110.
- [5] M. Aizenman and R. Fernández (1986). On the critical behavior of the magnetization in high dimensional Ising models. *J. Stat. Phys.*, **44**:393–454.
- [6] M. Aizenman and R. Graham (1983). On the renormalized coupling constant and the susceptibility in ϕ_4^4 field theory and the Ising model in four dimensions. *Nucl. Phys.*, **B225** [FS9]:261–288.
- [7] M. Aizenman, H. Kesten, and C.M. Newman (1987). Uniqueness of the infinite cluster and continuity of connectivity functions for short and long range percolation. *Commun. Math. Phys.*, **111**:505–531.
- [8] M. Aizenman and C.M. Newman (1984). Tree graph inequalities and critical behavior in percolation models. *J. Stat. Phys.*, **36**:107–143.
- [9] K.S. Alexander, J.T. Chayes, and L. Chayes (1990). The Wulff construction and asymptotics of the finite cluster distribution for two dimensional Bernoulli percolation. *Commun. Math. Phys.*, **131**:1–50.

- [10] Z. Alexandrowicz (1969). Monte Carlo of chains with excluded volume: a way to evade sample attrition. *J. Chem. Phys.*, **51**:561–565.
- [11] S.E. Alm (1992). Upper bounds for the connective constant of self-avoiding walks. Preprint.
- [12] S.E. Alm and S. Janson (1990). Random self-avoiding walks on one-dimensional lattices. *Commun. Statist.-Stochastic Models*, **6**:169–212.
- [13] D.J. Amit (1984). *Field Theory, the Renormalization Group, and Critical Phenomena*. World Scientific, Singapore, 2nd edition.
- [14] D.J. Amit, G. Parisi, and L. Peliti (1983). Asymptotic behaviour of the “true” self-avoiding walk. *Phys. Rev. B*, **27**:1635–1645.
- [15] C. Aragão de Carvalho and S. Caracciolo (1983). A new Monte Carlo approach to the critical properties of self-avoiding random walks. *J. Phys. (Paris)*, **44**:323–331.
- [16] C. Aragão de Carvalho, S. Caracciolo, and J. Fröhlich (1983). Polymers and $g|\phi|^4$ theory in four dimensions. *Nucl. Phys. B*, **215** [FS7]:209–248.
- [17] D. Arnaudon, D. Iagolnitzer, and J. Magnen (1991). Weakly self-avoiding polymers in four dimensions. Rigorous results. *Phys. Lett. B*, **273**:268–272.
- [18] M.N. Barber and B.W. Ninham (1970). *Random and Restricted Walks: Theory and Applications*. Gordon and Breach, New York.
- [19] D.J. Barsky and M. Aizenman (1991). Percolation critical exponents under the triangle condition. *Ann. Probab.*, **19**:1520–1536.
- [20] J. Batoulis and K. Kremer (1988). Statistical properties of biased sampling methods for long polymer chains. *J. Phys. A: Math. Gen.*, **21**:127–146.
- [21] B. Berg and D. Foerster (1981). Random paths and random surfaces on a digital computer. *Phys. Lett.*, **106B**:323–326.
- [22] J. van den Berg and H. Kesten (1985). Inequalities with applications to percolation and reliability. *J. Appl. Prob.*, **22**:556–569.
- [23] A. Berretti and A.D. Sokal (1985). New Monte Carlo method for the self-avoiding walk. *J. Stat. Phys.*, **40**:483–531.

- [24] P. Billingsley (1968). *Convergence of Probability Measures*. John Wiley and Sons, New York.
- [25] K. Binder and D.W. Heermann (1988). *Monte Carlo Simulation in Statistical Physics: An Introduction*. Springer-Verlag, New York.
- [26] E. Bolthausen (1990). On self-repellent one dimensional random walks. *Prob. Th. and Related Fields*, **86**:423–441.
- [27] E. Bolthausen (1992). On the construction of the three dimensional polymer measure. To appear in *Prob. Th. and Related Fields*.
- [28] J.-P. Bouchaud and A. Georges (1989). Flory formula as an extended law of large numbers. *Phys. Rev. B*, **39**:2846–2849.
- [29] A. Bovier, G. Felder, and J. Fröhlich (1984). On the critical properties of the Edwards and the self-avoiding walk model of polymer chains. *Nucl. Phys. B*, **230** [FS10]:119–147.
- [30] A. Bovier, J. Fröhlich, and U. Glaus (1986). Branched polymers and dimensional reduction. In K. Osterwalder and R. Stora, editors, *Critical Phenomena, Random Systems, Gauge Theories*. North-Holland, Amsterdam. (Les Houches 1984.)
- [31] R.G. Bowers and A. McKerrell (1973). An exact relation between the classical n -vector model ferromagnet and the self-avoiding walk problem. *J. Phys. C: Solid State Phys.*, **6**:2721–2732.
- [32] P. Bratley, B.L. Fox, and L.E. Schrage (1987). *A Guide to Simulation*. Springer-Verlag, New York, 2nd edition.
- [33] E. Brezin, J.C. Le Guillou, and J. Zinn-Justin (1973). Approach to scaling in renormalized perturbation theory. *Phys. Rev. D*, **8**:2418–2430.
- [34] P.J. Brockwell and R.A. Davis (1991). *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd edition.
- [35] D. Brydges, S.N. Evans, and J.Z. Imbrie (1992). Self-avoiding walk on a hierarchical lattice in four dimensions. *Ann. Probab.*, **20**:82–124.
- [36] D.C. Brydges (1986). A short course on cluster expansions. In K. Osterwalder and R. Stora, editors, *Critical Phenomena, Random Systems, Gauge Theories*. North-Holland, Amsterdam. (Les Houches 1984.)

- [37] D.C. Brydges, J. Fröhlich, and A.D. Sokal (1983). A new proof of the existence and nontriviality of the continuum φ_2^4 and φ_3^4 quantum field theories. *Commun. Math. Phys.*, **91**:141–186.
- [38] D.C. Brydges, J. Fröhlich, and T. Spencer (1982). The random walk representation of classical spin systems and correlation inequalities. *Commun. Math. Phys.*, **83**:123–150.
- [39] D.C. Brydges and T. Spencer (1985). Self-avoiding walk in 5 or more dimensions. *Commun. Math. Phys.*, **97**:125–148.
- [40] K. Burdzy and G.F. Lawler (1990a). Non-intersection exponents for random walk and Brownian motion. Part I. Existence and an invariance principle. *Prob. Th. and Rel. Fields*, **84**:393–410.
- [41] K. Burdzy and G.F. Lawler (1990b). Non-intersection exponents for random walk and Brownian motion. II. Estimates and application to a random fractal. *Ann. Probab.*, **18**:981–1009.
- [42] K. Burdzy, G.F. Lawler, and T. Polaski (1989). On the critical exponent for random walk intersections. *J. Stat. Phys.*, **56**:1–12.
- [43] M. Campanino, J.T. Chayes, and L. Chayes (1991). Gaussian fluctuations of connectivities in the subcritical regime of percolation. *Probab. Th. Rel. Fields*, **88**:269–341.
- [44] S. Caracciolo, A. Pelissetto, and A.D. Sokal (1990). Nonlocal Monte Carlo algorithm for self-avoiding walks with fixed endpoints. *J. Stat. Phys.*, **60**:1–53.
- [45] S. Caracciolo, A. Pelissetto, and A.D. Sokal (1992). Join-and-cut algorithm for self-avoiding walks with variable length and free endpoints. *J. Stat. Phys.*, **67**:65–111.
- [46] J.T. Chayes and L. Chayes (1986a). Percolation and random media. In K. Osterwalder and R. Stora, editors, *Critical Phenomena, Random Systems, Gauge Theories*. North-Holland, Amsterdam. (Les Houches 1984.)
- [47] J.T. Chayes and L. Chayes (1986b). Ornstein-Zernike behavior for self-avoiding walks at all noncritical temperatures. *Commun. Math. Phys.*, **105**:221–238.
- [48] J.T. Chayes and L. Chayes (1986c). Critical points and intermediate phases on wedges of Z^d . *J. Phys. A: Math. Gen.*, **19**:3033–3048.

- [49] L. Chayes (1991). On the critical behavior of the first passage time in $d \geq 3$. *Helv. Phys. Acta*, **64**:1055–1071.
- [50] J. Cheeger (1970). A lower bound for the lowest eigenvalue of the Laplacian. In R.C. Gunning, editor, *Problems in Analysis: A Symposium in Honor of S. Bochner*. Princeton Univ. Press, Princeton.
- [51] J. des Cloiseaux and G. Jannink (1990). *Polymers in Solution: Their Modelling and Structure*. Oxford University Press, New York.
- [52] D.R. Cox and D.V. Hinkley (1974). *Theoretical Statistics*. Chapman and Hall, London.
- [53] J.G. Curro (1974). Computer simulation of multiple chain systems—the effect of density on the average chain dimensions. *J. Chem. Phys.*, **61**:1203–1207.
- [54] P. Diaconis and D. Stroock (1991). Geometric bounds for eigenvalues of Markov chains. *Ann. Appl. Probab.*, **1**:36–61.
- [55] M. Doi and S.F. Edwards (1986). *The Theory of Polymer Dynamics*. Clarendon Press, Oxford.
- [56] C. Domb (1976). Finite cluster partition functions for the D -vector model. *J. Phys. A: Math. Gen.*, **9**:983–998.
- [57] C. Domb and M.E. Fisher (1958). On random walks with restricted reversals. *Proc. Camb. Phil. Soc.*, **54**:48–59.
- [58] C. Domb and G.S. Joyce (1972). Cluster expansion for a polymer chain. *J. Phys. C: Solid State Phys.*, **5**:956–976.
- [59] L.E. Dubins, A. Orłitsky, J.A. Reeds, and L.A. Shepp (1988). Self-avoiding random loops. *IEEE Trans. Inform. Theory*, **34**:1509–1516.
- [60] B. Duplantier (1988). Intersections of random walks. A direct renormalization approach. *Commun. Math. Phys.*, **117**:279–329.
- [61] B. Duplantier (1989). Fractals in two dimensions and conformal invariance. *Physica D*, **38**:71–87.
- [62] B. Duplantier (1990). Renormalization and conformal invariance for polymers. In H. van Beijeren, editor, *Fundamental Problems in Statistical Mechanics VII*, pages 171–223. Elsevier Science Publishers B.V., Amsterdam.

- [63] B. Duplantier and K.-H. Kwon (1988). Conformal invariance and intersections of random walks. *Phys. Rev. Lett.*, **61**:2514–2517.
- [64] R. Durrett (1988). *Lecture Notes on Particle Systems and Percolation*. Wadsworth & Brooks/Cole, Pacific Grove.
- [65] S.F. Edwards (1965). The statistical mechanics of polymers with excluded volume. *Proc. Phys. Soc. London*, **85**:613–624.
- [66] R.S. Ellis (1985). *Entropy, Large Deviations, and Statistical Mechanics*. Springer, Berlin.
- [67] M.H. Ernst (1988). Random walks with short memory. *J. Stat. Phys.*, **53**:191–201.
- [68] F. Family and M. Daoud (1984). Experimental realization of true self-avoiding walks. *Phys. Rev. B*, **29**:1506–1507.
- [69] G. Felder and J. Fröhlich (1985). Intersection probabilities of simple random walks: A renormalization group approach. *Commun. Math. Phys.*, **97**:111–124.
- [70] W. Feller (1968). *An Introduction to Probability Theory and Its Applications*, Volume I. Wiley, New York, third edition.
- [71] R. Fernández, J. Fröhlich, and A.D. Sokal (1992). *Random Walks, Critical Phenomena, and Triviality in Quantum Field Theory*. Springer, Berlin.
- [72] M.E. Fisher (1966). Shape of a self-avoiding walk or polymer chain. *J. Chem. Phys.*, **44**:616–622.
- [73] M.E. Fisher (1969). Discussion following “Statistics of long chains with repulsive interactions” by J. des Cloiseaux. *J. Phys. Soc. Japan*, **26**, suppl.:42–45.
- [74] M.E. Fisher and D.S. Gaunt (1964). Ising model and self-avoiding walks on hypercubical lattices and “high-density” expansions. *Phys. Rev.*, **133**:A224–A239.
- [75] M.E. Fisher and M.F. Sykes (1959). Excluded-volume problem and the Ising model of ferromagnetism. *Phys. Rev.*, **114**:45–58.
- [76] P.J. Flory (1949). The configuration of a real polymer chain. *J. Chem. Phys.*, **17**:303–310.

- [77] P.J. Flory (1971). *Principles of Polymer Chemistry*. Cornell University Press, Ithaca.
- [78] P.J. Flory (1976). Spatial configuration of macromolecular chains. In H. Markovitz and E.F. Casassa, editors, *Polymer Science: Achievements and Prospects*. Wiley Interscience, New York.
- [79] K.F. Freed (1981). Polymers as self-avoiding walks. *Ann. Probab.*, **9**:537–556.
- [80] J.J. Freire and A. Horta (1976). Mean reciprocal distances of short polymethylene chains. Calculation of the translational diffusion coefficient of *n*-alkanes. *J. Chem. Phys.*, **65**:4049–4054.
- [81] J. Fröhlich (1982). On the triviality of φ_d^4 theories and the approach to the critical point in $d \geq 4$ dimensions. *Nucl. Phys.*, **B200** [FS4]:281–296.
- [82] J. Fröhlich, B. Simon, and T. Spencer (1976). Infrared bounds, phase transitions, and continuous symmetry breaking. *Commun. Math. Phys.*, **50**:79–95.
- [83] P.J. Gans (1965). Self-avoiding random walks. I. Simple properties of intermediate-length walks. *J. Chem. Phys.*, **42**:4159–4163.
- [84] A. Gelman and D.B. Rubin (1992). Inference from iterative simulation using multiple sequences. To appear in *Statist. Sci.*
- [85] P.G. de Gennes (1972). Exponents for the excluded volume problem as derived by the Wilson method. *Phys. Lett.*, **A38**:339–340.
- [86] P.G. de Gennes (1979). *Scaling Concepts in Polymer Physics*. Cornell University Press, Ithaca.
- [87] C. Geyer (1992). Practical Markov chain Monte Carlo. To appear in *Statist. Sci.*
- [88] J. Glimm and A. Jaffe (1987). *Quantum Physics, A Functional Integral Point of View*. Springer, Berlin, 2nd edition.
- [89] S. Golowich and J.Z. Imbrie (1992). A new approach to the long-time behavior of self-avoiding random walks. *Ann. Phys.*, **217**:142–169.
- [90] P. Griffiths and J. Harris (1978). *Principles of Algebraic Geometry*. Wiley, New York.
- [91] G. Grimmett (1989). *Percolation*. Springer, Berlin.

- [92] R. Grishman (1973). Mean square endpoint separation of off-lattice self-avoiding walks. *J. Chem. Phys.*, **58**:220–225.
- [93] A.J. Guttmann (1978). On the zero-field susceptibility in the $d = 4$, $n = 0$ limit: analysing for confluent logarithmic singularities. *J. Phys. A: Math. Gen.*, **11**:L103–L106.
- [94] A.J. Guttmann (1981). Correction to scaling exponents and critical properties of the n -vector model with dimensionality > 4 . *J. Phys. A: Math. Gen.*, **14**:233–239.
- [95] A.J. Guttmann (1983). Bounds on connective constants for self-avoiding walks. *J. Phys. A: Math. Gen.*, **16**:2233–2238.
- [96] A.J. Guttmann (1987). On the critical behaviour of self-avoiding walks. *J. Phys. A: Math. Gen.*, **20**:1839–1854.
- [97] A.J. Guttmann (1989a). Asymptotic analysis of power-series expansions. In C. Domb and J.L. Lebowitz, editors, *Phase Transitions and Critical Phenomena*, Volume 11. Academic Press, New York.
- [98] A.J. Guttmann (1989b). On the critical behaviour of self-avoiding walks: II. *J. Phys. A: Math. Gen.*, **22**:2807–2813.
- [99] A.J. Guttmann and R.J. Bursill (1990). Critical exponents for the loop erased self-avoiding walk by Monte-Carlo methods. *J. Stat. Phys.*, **59**:1–9.
- [100] A.J. Guttmann and I.G. Enting (1988). The size and number of rings on the square lattice. *J. Phys. A: Math. Gen.*, **21**:L165–L172.
- [101] A.J. Guttmann and J. Wang (1991). The extension of self-avoiding random walk series in two dimensions. *J. Phys. A: Math. Gen.*, **24**:3107–3109.
- [102] A.J. Guttmann and N.C. Wormald (1984). On the number of spiral self-avoiding walks. *J. Phys. A: Math. Gen.*, **17**:L271–L274.
- [103] J.W. Halley and C. Dasgupta (1983). Percolation and related systems in equilibrium statistical mechanics. In B.D. Hughes and B.W. Ninham, editors, *The Mathematics and Physics of Disordered Media*, pages 260–282. Springer, Berlin. Lecture Notes in Mathematics 1035.
- [104] J.M. Hammersley (1961a). The number of polygons on a lattice. *Proc. Camb. Phil. Soc.*, **57**:516–523.

- [105] J.M. Hammersley (1961b). On the rate of convergence to the connective constant of the hypercubical lattice. *Quart. J. Math. Oxford*, (2), **12**:250–256.
- [106] J.M. Hammersley (1991). Self-avoiding walks. *Physica A*, **177**:51–57.
- [107] J.M. Hammersley (1992). Corrigendum: Self-avoiding walks. *Physica A*, **183**:574–578.
- [108] J.M. Hammersley and D.C. Handscomb (1964). *Monte Carlo Methods*. Chapman and Hall, London.
- [109] J.M. Hammersley and K.W. Morton (1954). Poor man's Monte Carlo. *J. Roy. Stat. Soc. B*, **16**:23–38.
- [110] J.M. Hammersley and D.J.A. Welsh (1962). Further results on the rate of convergence to the connective constant of the hypercubical lattice. *Quart. J. Math. Oxford*, (2), **13**:108–110.
- [111] J.M. Hammersley and S.G. Whittington (1985). Self-avoiding walks in wedges. *J. Phys. A: Math. Gen.*, (2), **18**:101–111.
- [112] T. Hara (1990). Mean field critical behaviour for correlation length for percolation in high dimensions. *Prob. Th. and Rel. Fields*, **86**:337–385.
- [113] T. Hara and G. Slade (1990a). Mean-field critical behaviour for percolation in high dimensions. *Commun. Math. Phys.*, **128**:333–391.
- [114] T. Hara and G. Slade (1990b). On the upper critical dimension of lattice trees and lattice animals. *J. Stat. Phys.*, **59**:1469–1510.
- [115] T. Hara and G. Slade (1992a). Self-avoiding walk in five or more dimensions. I. The critical behaviour. *Commun. Math. Phys.*, **147**:101–136.
- [116] T. Hara and G. Slade (1992b). The lace expansion for self-avoiding walk in five or more dimensions. *Reviews in Math. Phys.*, **4**:235–327.
- [117] T. Hara and G. Slade (1992c). The number and size of branched polymers in high dimensions. *J. Stat. Phys.*, **67**:1009–1038.
- [118] G.H. Hardy and S. Ramanujan (1917). Asymptotic formulae for the distribution of integers of various types. *Proc. Lond. Math. Soc.*, (2) **16**:112–132.

- [119] K. Hattori (1992). Self-avoiding processes on the Sierpinski gasket. In K. D. Elworthy and N. Ikeda, editors, *Asymptotic Problems in Probability Theory*. Longman, London, to appear.
- [120] O.J. Heilmann (1968). Mathematical polymers. I. *Mat. Fys. Medd. Dan. Vid. Selsk.*, **37**:3.
- [121] S. Hemmer and P.C. Hemmer (1984). An average self-avoiding random walk on the square lattice lasts 71 steps. *J. Chem. Phys.*, **81**:584–585.
- [122] C. Itzykson and J-M. Drouffe (1989). *Statistical Field Theory*, Volumes 1 and 2. Cambridge University Press, Cambridge.
- [123] E. J. Janse van Rensburg (1992a). Ergodicity of the BFACF algorithm in three dimensions. *J. Phys. A: Math. Gen.*, **25**:1031–1042.
- [124] E. J. Janse van Rensburg (1992b). On the number of trees in Z^d . *J. Phys. A: Math. Gen.*, **25**:3523–3528.
- [125] E. J. Janse van Rensburg and S.G. Whittington (1991). The BFACF algorithm and knotted polygons. *J. Phys. A: Math. Gen.*, **24**:5553–5567.
- [126] E. J. Janse van Rensburg, S.G. Whittington, and N. Madras (1990). The pivot algorithm and polygons: results on the FCC lattice. *J. Phys. A: Math. Gen.*, **23**:1589–1612.
- [127] P.C. Jurs and J.E. Reissner (1971). Efficient Monte Carlo generation of self-avoiding polymers. *J. Chem. Phys.*, **55**:4948–4951.
- [128] S. Karlin and H.M. Taylor (1975). *A First Course in Stochastic Processes*. Academic Press, New York, 2nd edition.
- [129] H. Kesten (1963). On the number of self-avoiding walks. *J. Math. Phys.*, **4**:960–969.
- [130] H. Kesten (1964). On the number of self-avoiding walks. II. *J. Math. Phys.*, **5**:1128–1137.
- [131] H. Kesten (1982). *Percolation Theory for Mathematicians*. Birkhäuser, Boston.
- [132] H. Kesten (1987). Percolation theory and first passage percolation. *Ann. Probab.*, **15**:1231–1271.

- [133] H. Kesten (1990). Asymptotics in high dimensions for percolation. In G.R. Grimmett and D.J.A. Welsh, editors, *Disorder in Physical Systems*. Clarendon Press, Oxford.
- [134] J.F.C. Kingman (1972). *Regenerative Phenomena*. Wiley, New York.
- [135] C. Kipnis and S.R.S. Varadhan (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.*, **104**:1–19.
- [136] D.A. Klarner (1967). Cell growth problems. *Canad. J. Math.*, **19**:851–863.
- [137] D.J. Klein (1980). Asymptotic distributions for self-avoiding walks constrained to strips, cylinders, and tubes. *J. Stat. Phys.*, **23**:561–586.
- [138] D.J. Klein (1981). Rigorous results for branched polymer models with excluded volume. *J. Chem. Phys.*, **75**:5186–5189.
- [139] D.E. Knuth (1973). *The Art of Computer Programming*, Volume 3. Addison Wesley, Reading.
- [140] K. Kremer and K. Binder (1988). Monte Carlo simulation of lattice models for macromolecules. *Comp. Phys. Rep.*, **7**:259–310.
- [141] A.K. Kron (1965). Monte-Carlo method in the statistical calculations of macromolecules. *Vysokomolekul. Soedin. [Polymer Sci. USSR]*, **7**:1228–1234.
- [142] A.K. Kron, O.B. Ptitsyn, A.M. Skvortsov, and A.K. Federov (1967). A study of statistical globula-coil transition in macromolecules using the Monte-Carlo technique. *Molek. Biol.*, **1**:576–582.
- [143] M. Lal (1969). 'Monte Carlo' computer simulations of chain molecules. I. *Molec. Phys.*, **17**:57–64.
- [144] A.I. Larkin and D.E. Khmel'Nitskii (1969). Phase transition in uniaxial ferroelectrics. *Soviet Physics JETP*, **29**:1123–1128. English translation of *Zh. Eksp. Teor. Fiz.* **56**, 2087–2098, (1969).
- [145] G.F. Lawler (1980). A self-avoiding random walk. *Duke Math. J.*, **47**:655–693.
- [146] G.F. Lawler (1982). The probability of intersection of independent random walks in four dimensions. *Commun. Math. Phys.*, **86**:539–554.

- [147] G.F. Lawler (1985a). Intersections of random walks in four dimensions. II. *Commun. Math. Phys.*, **97**:583–594.
- [148] G.F. Lawler (1985b). The probability of intersection of three random walks in three dimensions. Unpublished manuscript.
- [149] G.F. Lawler (1986). Gaussian behavior of loop-erased self-avoiding random walk in four dimensions. *Duke Math. J.*, **53**:249–269.
- [150] G.F. Lawler (1988). Loop-erased self-avoiding random walk in two and three dimensions. *J. Stat. Phys.*, **50**:91–108.
- [151] G.F. Lawler (1989). The infinite self-avoiding walk in high dimensions. *Ann. Probab.*, **17**:1367–1376.
- [152] G.F. Lawler (1991). *Intersections of Random Walks*. Birkhäuser, Boston.
- [153] G.F. Lawler and A.D. Sokal (1988). Bounds on the L^2 spectrum for Markov chains and Markov processes: A generalization of Cheeger's inequality. *Trans. Amer. Math. Soc.*, **309**:557–580.
- [154] J.C. Le Guillou and J. Zinn-Justin (1989). Accurate critical exponents from field theory. *J. Phys. France*, **50**:1365–1370.
- [155] B. Li and A.D. Sokal (1990). High-precision Monte-Carlo test of the conformal-invariance predictions for two-dimensional mutually avoiding walks. *J. Stat. Phys.*, **61**:723–748.
- [156] E.H. Lieb (1980). A refinement of Simon's correlation inequality. *Commun. Math. Phys.*, **77**:127–136.
- [157] T.C. Lubensky and J. Isaacson (1979). Statistics of lattice animals and dilute branched polymers. *Phys. Rev.*, **A20**:2130–2146.
- [158] B. MacDonald, N. Jan, D.L. Hunter, and M.O. Steinitz (1985). Polymer conformations through 'wiggling'. *J. Phys. A: Math. Gen.*, **18**:2627–2631.
- [159] N. Madras (1988). End patterns of self-avoiding walks. *J. Stat. Phys.*, **53**:689–701.
- [160] N. Madras (1991a). On the Ornstein-Zernike decay of self-avoiding walks. In G.J. Morrow and W.-S. Yang, editors, *Proceedings of the Conference on Probability Models in Mathematical Physics*. World Scientific, Singapore.

- [161] N. Madras (1991b). Bounds on the critical exponent of self-avoiding polygons. In R. Durrett and H. Kesten, editors, *Random Walks, Brownian Motion and Interacting Particle Systems*. Birkhäuser, Boston.
- [162] N. Madras, A. Orłitsky, and L.A. Shepp (1990). Monte Carlo generation of self-avoiding walks with fixed endpoints and fixed length. *J. Stat. Phys.*, **58**:159–183.
- [163] N. Madras and A.D. Sokal (1987). Nonergodicity of local, length-conserving Monte Carlo algorithms for the self-avoiding walk. *J. Stat. Phys.*, **47**:573–595.
- [164] N. Madras and A.D. Sokal (1988). The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *J. Stat. Phys.*, **50**:109–186.
- [165] F. Mandel (1979). Macromolecular dimensions obtained by an efficient Monte Carlo method: The mean square end-to-end separation. *J. Chem. Phys.*, **70**:3984–3988.
- [166] B. Masand, U. Wilensky, J.P. Massar, and S. Redner (1992). An extension of the two-dimensional self-avoiding walk series on the square lattice. *J. Phys. A: Math. Gen.*, **25**:L365–L369.
- [167] F.L. McCrackin (1972). Weighting methods for Monte Carlo calculation of polymer configurations. *J. Res. Nat. Bur. Stand.—B. Math. Sci.*, **76B**:193–200.
- [168] E.W. Montroll and B.J. West (1979). On an enriched collection of stochastic processes. In E.W. Montroll and J.L. Lebowitz, editors, *Studies in Statistical Mechanics, Volume VII: Fluctuation Phenomena*. North Holland, Amsterdam.
- [169] B.G. Nguyen (1987). Gap exponents for percolation processes with triangle condition. *J. Stat. Phys.*, **49**:235–243.
- [170] B.G. Nguyen and W.-S. Yang (1991). Triangle condition for oriented percolation in high dimensions. Preprint.
- [171] B. Nienhuis (1982). Exact critical exponents of the $O(n)$ models in two dimensions. *Phys. Rev. Lett.*, **49**:1062–1065.
- [172] B. Nienhuis (1984). Critical behavior of two-dimensional spin models and charge asymmetry in the Coulomb gas. *J. Stat. Phys.*, **34**:731–761.

- [173] B. Nienhuis (1987). Coulomb gas formulation of two-dimensional phase transitions. In C. Domb and J.L. Lebowitz, editors, *Phase Transitions and Critical Phenomena*, Volume 11. Academic Press, New York.
- [174] E. Nummelin (1984). *General Irreducible Markov Chains and Non-negative Operators*. Cambridge University Press, Cambridge.
- [175] G.L. O'Brien (1990). Monotonicity of the number of self-avoiding walks. *J. Stat. Phys.*, **59**:969–979.
- [176] S.P. Obukhov (1980). The problem of directed percolation. *Physica*, **101A**:145–155.
- [177] O.F. Olaj and K.H. Pelinka (1976). Pair distribution function and pair potential of lattice models under theta condition. I. Numerical evaluation. *Makromol. Chem.*, **177**:3413–3425.
- [178] S. Orey (1971). *Limit Theorems for Markov Chain Transition Probabilities*. Van Nostrand Reinhold, London.
- [179] L.S. Ornstein and F. Zernike (1914). Accidental deviations of density and opalescence at the critical point of a single substance. *K. Akad. Amsterdam*, **17**:793–806.
- [180] G. Parisi (1988). *Statistical Field Theory*. Addison Wesley, New York.
- [181] Y. M. Park (1989). Direct estimates on intersection probabilities of random walks. *J. Stat. Phys.*, **57**:319–331.
- [182] L. Peliti and L. Pietronero (1987). Random walks with memory. *Rivista del Nuovo Cimento*, **10**:1–33.
- [183] M.D. Penrose (1992). On the spread-out limit for bond and continuum percolation. To appear in *Ann. Appl. Probab.*
- [184] P.H. Peskun (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, **60**:607–612.
- [185] L. Pietronero (1983). Critical dimensionality and exponent of the "true" self-avoiding walk. *Phys. Rev. B*, **27**:5887–5889.
- [186] N. Pippenger (1989). Knots in random walks. *Discrete Appl. Math.*, **25**:273–278.
- [187] M.B. Priestley (1981). *Spectral Analysis and Time Series*. Academic Press, London.

- [188] D.C. Rapaport (1985). On three-dimensional self-avoiding walks. *J. Phys. A: Math. Gen.*, **18**:113–126.
- [189] S. Redner and P.J. Reynolds (1981). Position-space renormalisation group for isolated polymer chains. *J. Phys. A: Math. Gen.*, **14**:2679–2703.
- [190] M. Reed and B. Simon (1972). *Methods of Modern Mathematical Physics. I: Functional Analysis*. Academic Press, New York.
- [191] J. Reiter (1990). Monte Carlo simulations of linear and cyclic chains on cubic and quadratic lattices. *Macromolecules*, **23**:3811–3816.
- [192] M.N. Rosenbluth and A.W. Rosenbluth (1955). Monte Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.*, **23**:356–359.
- [193] D. Ruelle (1969). *Statistical Mechanics, Rigorous Results*. W.A. Benjamin, Inc., New York.
- [194] S.S. Shapiro and M.B. Wilk (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52**:591–611.
- [195] Z. Šidák (1964). Eigenvalues of operators in l_p -spaces of denumerable Markov chains. *Czech. Math. J.*, **14**:438–443.
- [196] S.D. Silvey (1975). *Statistical Inference*. Chapman and Hall, London.
- [197] B. Simon (1980). Correlation inequalities and the decay of correlations in ferromagnets. *Commun. Math. Phys.*, **77**:111–126.
- [198] G. Slade (1987). The diffusion of self-avoiding random walk in high dimensions. *Commun. Math. Phys.*, **110**:661–683.
- [199] G. Slade (1988). Convergence of self-avoiding random walk to Brownian motion in high dimensions. *J. Phys. A: Math. Gen.*, **21**:L417–L420.
- [200] G. Slade (1989). The scaling limit of self-avoiding random walk in high dimensions. *Ann. Probab.*, **17**:91–107.
- [201] G. Slade (1991). The lace expansion and the upper critical dimension for percolation. *Lectures in Applied Mathematics*, **27**:53–63. (Mathematics of Random Media, eds. W.E. Kohler and B.S. White, A.M.S., Providence).

- [202] A.D. Sokal (1979). A rigorous inequality for the specific heat of an Ising or φ^4 ferromagnet. *Phys. Lett.*, **71A**:451–453.
- [203] A.D. Sokal (1982). An alternate constructive approach to the φ_4^4 quantum field theory, and a possible destructive approach to φ_4^4 . *Ann. Inst. Henri Poincaré*, **37**:317–398.
- [204] A.D. Sokal (1989). Monte Carlo methods in statistical mechanics: Foundations and new algorithms. Lecture notes: Cours de Troisième Cycle de la Physique en Suisse Romande (Lausanne, June 1989).
- [205] A.D. Sokal (1991). How to beat critical slowing-down: 1990 update. *Nucl. Phys. B (Proc. Suppl.)*, **20**:55–67.
- [206] A.D. Sokal and L.E. Thomas (1988). Absence of mass gap for a class of stochastic contour models. *J. Stat. Phys.*, **51**:907–947.
- [207] A.D. Sokal and L.E. Thomas (1989). Exponential convergence to equilibrium for a class of random walk models. *J. Stat. Phys.*, **54**:797–828.
- [208] C.E. Soteros, D.W. Sumners, and S.G. Whittington (1992). Entanglement complexity of graphs in Z^3 . *Math. Proc. Camb. Phil. Soc.*, **111**:75–91.
- [209] C.E. Soteros and S.G. Whittington (1988). Polygons and stars in a slit geometry. *J. Phys. A: Math. Gen.*, **21**:L857–L861.
- [210] C.E. Soteros and S.G. Whittington (1989). Lattice models of branched polymers: effects of geometrical constraints. *J. Phys. A: Math. Gen.*, **22**:5259–5270.
- [211] F. Spitzer (1976). *Principles of Random Walk*. Springer, New York, 2nd edition.
- [212] A.J. Stam (1971). Renewal theory in r dimensions II. *Comp. Math.*, **23**:1–13.
- [213] D. Stauffer (1985). *Introduction to Percolation Theory*. Taylor and Francis, London.
- [214] S.D. Stellman and P.J. Gans (1972). Efficient computer simulation of polymer conformation. I. Geometric properties of the hard-sphere model. *Macromolecules*, **5**:516–526.
- [215] A. Stoll (1989). Invariance principles for Brownian intersection local time and polymer measures. *Math. Scand.*, **64**:133–160.

- [216] D.W. Sumners and S.G. Whittington (1988). Knots in self-avoiding walks. *J. Phys. A: Math. Gen.*, **21**:1689–1694.
- [217] K. Suzuki (1968). The excluded volume effect of very-long-chain molecules. *Bull. Chem. Soc. Japan*, **41**:538.
- [218] M.F. Sykes, D.S. McKenzie, M.G. Watts, and J.L. Martin (1972). The number of self-avoiding rings on a lattice. *J. Phys. A: Gen. Phys.*, **5**:661–666.
- [219] K. Symanzik (1969). Euclidean quantum field theory. In R. Jost, editor, *Local Quantum Field Theory*. Academic Press, New York.
- [220] H. Tasaki and T. Hara (1987). Critical behaviour in a system of branched polymers. *Prog. Theor. Phys. Suppl.*, **92**:14–25.
- [221] C.J. Thompson (1988). *Classical Equilibrium Statistical Mechanics*. Oxford University Press, Oxford.
- [222] S.R.S. Varadhan (1969). Appendix to: Euclidean quantum field theory, by K. Symanzik. In R. Jost, editor, *Local Quantum Field Theory*. Academic Press, New York.
- [223] P.H. Verdier (1969). A simulation model for the study of the motion of random-coil polymer chains. *J. Comput. Phys.*, **4**:204–210.
- [224] P.H. Verdier and W.H. Stockmayer (1962). Monte Carlo calculations on the dynamics of polymers in dilute solution. *J. Chem. Phys.*, **36**:227–235.
- [225] F.T. Wall and J.J. Erpenbeck (1959). New method for the statistical computation of polymer dimensions. *J. Chem. Phys.*, **30**:634–637.
- [226] F.T. Wall and F. Mandel (1975). Macromolecular dimensions obtained by an efficient Monte Carlo method without sample attrition. *J. Chem. Phys.*, **63**:4592–4595.
- [227] G.S. Watson (1983). *Statistics on Spheres*. John Wiley and Sons, New York.
- [228] F.J. Wegner and E.K. Riedel (1973). Logarithmic corrections to the molecular-field behavior of critical and tricritical systems. *Phys. Rev. B*, **7**:248–256.
- [229] J. Westwater (1980). On Edwards' model for long polymer chains. *Commun. Math. Phys.*, **72**:131–174.

- [230] J. Westwater (1982). On Edwards' model for long polymer chains III. Borel summability. *Commun. Math. Phys.*, **84**:459–470.
- [231] J. Westwater (1985). On Edwards' model for long polymer chains. In S. Albevario and P. Blanchard, editors, *Trends and Developments in the Eighties. Bielefeld Encounters in Mathematical Physics IV/V*. World Scientific, Singapore.
- [232] S.G. Whittington (1982). Statistical mechanics of polymer solutions and polymer adsorption. *Adv. Chem. Phys.*, **51**:1–48.
- [233] S.G. Whittington (1983). Self-avoiding walks with geometrical constraints. *J. Stat. Phys.*, **30**:449–456.
- [234] S.G. Whittington and C.E. Soteris (1991). Polymers in slabs, slits, and pores. *Israel J. Chem.*, **31**:127–133.
- [235] W.-S. Yang and D. Klein (1988). A note on the critical dimension for weakly self-avoiding walks. *Prob. Th. and Rel. Fields*, **79**:99–114.
- [236] W.-S. Yang and Y. Zhang (1992). A note on differentiability of the cluster density for independent percolation in high dimensions. *J. Stat. Phys.*, **66**:1123–1138.
- [237] K. Yosida (1980). *Functional Analysis*. Springer-Verlag, Berlin, 6th edition.

Notation

$$\begin{aligned} \|x\|_p &= (\sum_{\mu=1}^d |x|_\mu^p)^{1/p} \\ |x| &= \|x\|_2 \\ \|x\|_\infty &= \max\{|x_\mu| : \mu = 1, \dots, d\} \\ x \cdot y &= \sum_{\mu=1}^d x_\mu y_\mu \end{aligned}$$

$$\begin{aligned} \partial_z &= \frac{\partial}{\partial z} \\ \partial_\mu &= \frac{\partial}{\partial k_\mu} \\ \nabla_k &= (\partial_1, \dots, \partial_d) \\ \nabla_k^2 &= \sum_{\mu=1}^d \partial_\mu^2 \end{aligned}$$

$$\begin{aligned} \hat{f}(k) &= \sum_{x \in \mathbb{Z}^d} f(x) e^{ik \cdot x} \\ f(x) &= (2\pi)^{-d} \int_{[-\pi, \pi]^d} \hat{f}(k) e^{-ik \cdot x} d^d k \\ \|f\|_p &= [\sum_{x \in \mathbb{Z}^d} |f(x)|^p]^{1/p} \\ \|f\|_\infty &= \sup_{x \in \mathbb{Z}^d} |f(x)| \\ \|\hat{f}\|_p &= [(2\pi)^{-d} \int_{[-\pi, \pi]^d} |\hat{f}(k)|^p d^d k]^{1/p} \\ f * g(x) &= \sum_{y \in \mathbb{Z}^d} f(x-y)g(y) \end{aligned}$$

$\bar{z}$	complex conjugate of z
Re	real part of a complex number
$\text{Res}(f(z), z_0)$	residue of $f(z)$ at z_0
$[x]$	greatest integer less than or equal to x
$ S $	cardinality of the set S
$f(x) \sim g(x)$	$\lim f(x)/g(x) = 1$
$f(x) \simeq g(x)$	existence of c_1, c_2 such that $c_1 g(x) \leq f(x) \leq c_2 g(x)$
$f(x) \approx g(x)$	unproven belief that $f(x)$ and $g(x)$ have similar asymptotic behaviour
$f(x) = O(g(x))$	existence of K such that $ f(x) \leq K g(x) $
$f(x) = o(g(x))$	$\lim f(x)/g(x) = 0$