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in five dimensions, and to Takashi Hara for the pleasure of four years of collaboration and for permission to include unpublished joint work.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** $\mathbf{Z}^d$. The points of this graph are the points of the d-dimensional Euclidean space $\mathbf{R}^d$ 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: $x = (x_1, x_2, \ldots, x_d)$. The usual Euclidean dot product on $\mathbf{Z}^d$ 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 $||x||_\infty = \max\{|x_i|: i = 1, \ldots, d\}$.

An N-step self-avoiding walk ω on $\mathbf{Z}^d$, beginning at the site x, is defined as a sequence of sites $(\omega(0), \omega(1), \ldots, \omega(N))$ 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 $\omega_i(j)$ $(i = 1, \ldots, d)$. 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

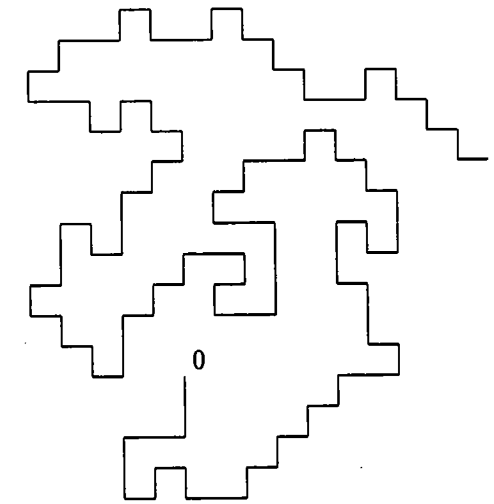


Figure 1.1: A two-dimensional self-avoiding walk with 115 steps.

for the asymptotic form of c_N as $N \to \infty$. 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

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

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*

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

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 $\mathbf{Z}^d$, starting at the origin, is a sequence $\omega = (\omega(0), \omega(1), \ldots, \omega(N))$ 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 $S_N = X^{(1)} + X^{(2)} + \ldots + X^{(N)}$. Expanding $|S_N|^2$, the mean-square displacement is given by

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

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 $\mathbf{Z}^d$ in which steps lie in a symmetric finite set $\Omega \subset \mathbf{Z}^d$ 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

$$c_N \sim A\mu^N N^{\gamma-1} \tag{1.1.4}$$

and

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

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 $N \to \infty$:

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

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

$$\mu = \lim_{N\to\infty} c_N^{1/N} \tag{1.1.6}$$

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

$$d \leq \mu \leq 2d - 1. \tag{1.1.7}$$

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 $d \to \infty$

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

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

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

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

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

If $\gamma > 1$ then this probability goes to zero as $N \to \infty$, while if $\gamma = 1$ it remains positive. For the simple random walk the analogous probability is known to remain positive as $N \to \infty$ 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 $|\omega(N)|^2$ 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 $N \to \infty$ for $d > 4$, but not for $d \leq 4$.

The conjectured values of γ and ν are as follows:

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

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

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:

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

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

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 $\nu_{Flory} = 3/(2+d)$ for $d \le 4$ and $\nu_{Flory} = 1/2$ for $d > 4$. This agrees with Equation (1.1.12) for $d = 2$ and $d \ge 4$ (apart from the logarithmic correction when $d = 4$), and comes very close for $d = 3$. The exact Flory value $\nu_{Flory} = 3/5$ 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

$$\mu = \lim_{N\to\infty} c_N^{1/N} \tag{1.2.1}$$

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

$$d \le \mu \le 2d - 1. \tag{1.2.2}$$

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

[1] This 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.

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*

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

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

$$c_{N+M} \le c_N c_M. \tag{1.2.3}$$

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*:

$$\log c_{N+M} \le \log c_N + \log c_M. \tag{1.2.4}$$

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 $\{a_n\}_{n\ge1}$ be a sequence of real numbers which is subadditive, i.e., $a_{n+m} \le a_n + a_m$. Then the limit $\lim_{n\to\infty} n^{-1}a_n$ exists in $[-\infty, \infty)$ and is equal to*

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\ge1} \frac{a_n}{n}. \tag{1.2.5}$$

Proof. It suffices to show that

$$\limsup_{n\to\infty} \frac{a_n}{n} \le \frac{a_k}{k} \tag{1.2.6}$$

for every k, since taking the $\liminf_{k\to\infty}$ in (1.2.6) gives existence of the limit, and then (1.2.5) can be seen by taking the $\inf_{k\ge1}$ in (1.2.6).

To prove (1.2.6), we fix k and let

$$A_k = \max_{1\le r\le k} a_r. \tag{1.2.7}$$

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 $1 \le r \le k$. Using subadditivity, we have

$$a_n \le ja_k + a_r \le \frac{n}{k}a_k + A_k. \tag{1.2.8}$$

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

Equation (1.2.5) shows that $\lim_{n\to\infty} n^{-1}a_n < \infty$. 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 $a_n \geq 0$. □

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

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

and hence

$$\mu^N \leq c_N, \quad N \geq 1. \tag{1.2.10}$$

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

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

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

$$\mu_\tau = \lim_{N\to\infty} c_{N,\tau}^{1/N} = \inf_{N\geq 1} c_{N,\tau}^{1/N}. \tag{1.2.12}$$

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 $\tau \to \infty$.*

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

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

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

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

$$\theta^3 - 2(d-1)\theta^2 - 2(d-1)\theta - 1 = 0. \tag{1.2.14}$$

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

$$b_M b_N \leq b_{M+N}. \tag{1.2.15}$$

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

$$\mu_{Bridge} \equiv \lim_{n\to\infty} b_n^{1/n} = \sup_{n\geq 1} b_n^{1/n} \tag{1.2.16}$$

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

$$b_N^{1/N} \leq \mu_{Bridge} \leq \mu. \tag{1.2.17}$$

In Section 3.1 it will be shown that in fact $\mu_{Bridge} = \mu$. 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 ± 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

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

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.,

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

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 $1 - 2dz$:

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

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

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

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

$$z_c \equiv \left[\lim_{N\to\infty} c_N^{1/N}\right]^{-1} = \frac{1}{\mu}, \tag{1.3.5}$$

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

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

and hence χ is "continuous" at the critical point, in the sense that $\chi(z) \to \infty$ 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

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

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

$$C_1 g(x) \leq f(x) \leq C_2 g(x) \tag{1.3.8}$$

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

$$c_N \simeq \mu^N N^{\gamma-1} \quad \text{as } N \to \infty, \tag{1.3.9}$$

it can be concluded that

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

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

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

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

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

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

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

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

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

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

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

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:

$$m(z) = \liminf_{n\to\infty} \frac{-\log G_z(0,(n,0,\ldots,0))}{n}. \tag{1.3.15}$$

In Theorem 4.1.3 it will be shown that in fact the lim inf 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 $(n,0,\ldots,0)$ gives

$$G_z(0,(n,0,\ldots,0)) \geq z^n, \tag{1.3.16}$$

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

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

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

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

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

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

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

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

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:

$$\begin{aligned}c_N &\sim A\mu^N N^{\gamma-1} && (1.4.1)\\ \chi(z) &\sim A'(z_c - z)^{-\bar{\gamma}} && (1.4.2)\\ \langle|\omega(N)|^2\rangle &\sim DN^{2\nu} && (1.4.3)\\ \xi(z) &\sim B(z_c - z)^{-\bar{\nu}} && (1.4.4)\\ \xi_p(z) &\sim B_p(z_c - z)^{-\nu_p}. && (1.4.5)\end{aligned}$$

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 $m(z) \to 0$ 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

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

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

$$\gamma = (2-\eta)\nu. \tag{1.4.9}$$

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

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

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

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

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*

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

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 $N \to \infty$ 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 $N \to \infty$ 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 $N \to \infty$ 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 $\mathbf{Z}^d$, there exist positive constants A_1 and A_2 and an integer N_0 (all depending on x and y) such that*

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

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

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

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 const.$N^{-d/2}\exp[-d|x|^2/2N] \sim$ const.$N^{-d/2}$, as $N \to \infty$. 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

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

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 $j = 0, \ldots, n$, we obtain $c_{N,N} \geq (N+1)c_N^2$. Together these bounds give

$$\frac{\gamma+1}{2} \leq \Delta_4 \leq \frac{\gamma+2}{2}. \tag{1.4.25}$$

This can be rewritten as

$$1 \leq 2\Delta_4 - \gamma \leq 2. \tag{1.4.26}$$

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

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

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

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

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

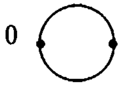
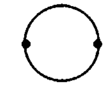
$$\chi(z) \simeq (z_c - z)^{-1} \quad \text{as } z \nearrow z_c. \tag{1.5.3}$$

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*

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

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

$\mathsf{B}(z) = \sum_x$ 0  x = 

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

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

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

$$\mathsf{B}(z_c) < \infty.$$

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*

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

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

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

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

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

$$= \sum_{x,y} \sum_{\substack{\omega^{(1)}:0\to x \\ \omega^{(2)}:x\to y}} z^{|\omega^{(1)}|+|\omega^{(2)}|} I[\omega^{(1)} \cap \omega^{(2)} = \{x\}]$$

$$\equiv Q(z), \tag{1.5.8}$$

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

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

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

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

Here the factor $\mathsf{B}(z) - 1$ arises from the two paths joining w and x. The upper bound involves $\mathsf{B}(z) - 1$ rather than $\mathsf{B}(z)$ 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

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

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

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

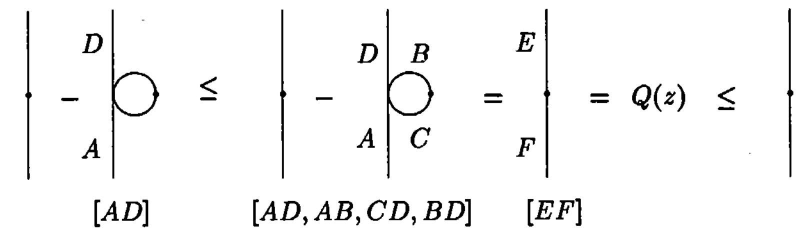


Figure 1.2: A diagrammatic representation of the inequality $\chi(z)^2 - Q(z)[\mathsf{B}(z) - 1] \leq Q(z) \leq \chi(z)^2$ occurring in the proof of Lemma 1.5.2. The list of pairs of lines indicates interactions between the propagators, in the sense that the corresponding walks must avoid each other.

Combining this inequality with (1.5.9) completes the proof of the lemma. □

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 $[\mathsf{B}(z)^{-1}, 1]$, 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*

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

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

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 $z \in [z_1, z_c)$

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

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

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

Rewriting gives

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

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. □

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

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

Using the fact that

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

a routine calculation using (1.5.16) gives the bound

$$\mathsf{B}(z) \leq C[1 + \log\chi(z)]. \qquad (1.5.17)$$

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

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

and therefore

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

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

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

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

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

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. □

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

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

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 $\mathsf{R}(z) < \mathsf{B}(z)$, which is defined by taking only those contributions to the bubble from pairs of walks which are mutually avoiding apart from their common endpoints:

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

Theorem 1.5.5 *If $\mathsf{B}(z_c) < \infty$ and in addition $\mathsf{R}(z_c) - 1 < 1/4$, 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 $\mathsf{R}(z_c) - 1$ in five dimensions is $0.434636 > 0.25$ [Hara and Slade (1991b)]. For $d = 6$ the same reference reports $\mathsf{B}(z_c) - 1 \leq 0.25974$. However the repulsive bubble in six dimensions satisfies $\mathsf{R}(z_c) - 1 \leq 0.2343 < 0.25$, 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, $\mathsf{B}(z_c) - 1 < 1/4$ 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

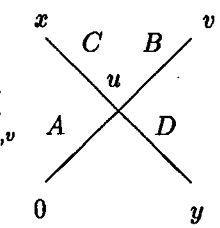
$$\sum_{N_1, N_2 = 0}^{\infty} c_{N_1,N_2} z^{N_1 + N_2} \simeq (z_c - z)^{-4}. \quad (1.5.24)$$

The left side is equal to

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

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):

$$\sum_{u,x,y,v}$$



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

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

$$\sum_{u,v,x,y} \sum_{\substack{\omega^{(1)}: 0 \to u \\ \omega^{(2)}: u \to v}} \sum_{\substack{\rho^{(1)}: x \to u \\ \rho^{(2)}: u \to y}} z^{|\omega^{(1)}| + |\omega^{(2)}| + |\rho^{(1)}| + |\rho^{(2)}|}$$
$$\times \Big\{ 1 - I[\omega^{(1)} \cap \omega^{(2)} \neq \{u\}] - I[\omega^{(1)} \cap \rho^{(1)} \neq \{u\}]$$
$$- I[\omega^{(1)} \cap \rho^{(2)} \neq \{u\}] - I[\rho^{(1)} \cap \rho^{(2)} \neq \{u\}] \Big\}. \quad (1.5.27)$$

This bound is equal to

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

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

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

for (1.5.26), assuming that $\mathsf{R}(z_c) - 1 < 1/4$. □

1.6 Notes

Section 1.1. Existence of the connective constant $\mu = \lim_{N\to\infty} c_N^{1/N}$ 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 $f(m,r) \leq \sum G^r n_1^H n_2^H \cdots n_r^H \mu^m$, where the sum is over all $n_1, \ldots, n_r \geq 1$ that sum to m. By the arithmetic-geometric inequality we have $n_1 n_2 \cdots n_r \leq (m/r)^r$, which implies $f(m,r) \leq \binom{m-1}{r-1} G^r (m/r)^{rH}\mu^m$. 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

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

(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 Glaus (1986) and Adler (1984) respectively.

For $d \geq 5$ it has been proven that $\hat{G}_{z_c}(k) \sim \text{const.}k^{-2}$ as $k \to 0$, 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 $g(x) = (2\pi)^{-d}\int_{T^d}\hat{g}(k)e^{-ik\cdot x}d^dk$. Suppose that there is a neighourhood $U \subset \mathbf{T}^d$ of 0 such that*

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

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

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

The following shows that in general the hypothesis of existence of $d-2$ derivatives for $\hat{h}$ cannot be relaxed: we give an example[2] of a function $\hat{g}$ on $\mathbf{T}^d$, for $d \geq 3$, which is asymptotic to a multiple of k^{-2} as $k \to 0$, 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 $k \to 0$. Fix q such that $d-3 < q < d-2$, and for $-\pi \leq t \leq \pi$ define

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

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

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

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

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

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

and

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

Then $\hat{g}(k)$ is asymptotic to $(2d)k^{-2}$ as $k \to 0$, 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 Glaus (1986), Tasaki and Hara (1987) and Hara and Slade (1990b).

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*

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

The proof relies on bridges (see Definition 1.2.4), and yields as a bonus the fact that $\mu_{Bridge} = \mu$. 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*

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

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 $A = A_1 + \cdots + A_k$ where $A_1 > \ldots > A_k$). Then*

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

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 $N \geq 1$,*

$$h_N \leq P_D(N) b_N, \tag{3.1.3}$$

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 $j = 1, 2, \ldots$, 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

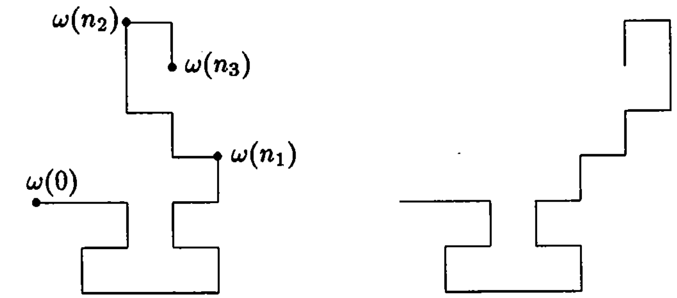


Figure 3.1: A half-space walk ω in $H_{20}[4, 2, 1]$ and the transformed walk ω' in $H_{20}[6, 1]$.

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

For every decreasing sequence of k positive integers $a_1 > a_2 \ldots > a_k > 0$, let $H_N[a_1, \ldots, a_k]$ be the set of N-step half-space walks ω with $\omega(0) = 0$ and $A_1(\omega) = a_1, \cdots, A_k(\omega) = a_k$, 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 $H_N[a_1, a_2, \cdots, a_k]$, then ω' is in $H_N[a_1 + a_2, a_3, \cdots, a_k]$; moreover, this transformation is one-to-one, so

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

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

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

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

which tells us that

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

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

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

which proves the proposition. □

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

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

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 $(\omega(m), \ldots, \omega(n))$ is a half-space walk, as is

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

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

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

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

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

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

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

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

In particular,

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

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. □

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,

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

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 $A_1 > A_2 > \ldots$ and $\sum m_i = N$; moreover, the sequence of bridges uniquely determines the original half-space walk. Therefore for $N \geq 1$

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

where the sum is over all integers $k \geq 1$, all integers $A_1 > \ldots > A_k > 0$, and all integers $m_1, \ldots, m_k \geq 1$ 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. □

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 $\omega(i)$ $(i = 1, \ldots, N-1)$.*

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, 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 $\mathbf{Z}^2$, 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 $\mathbf{R}^d$ 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 $\mathbf{Z}^d$, 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

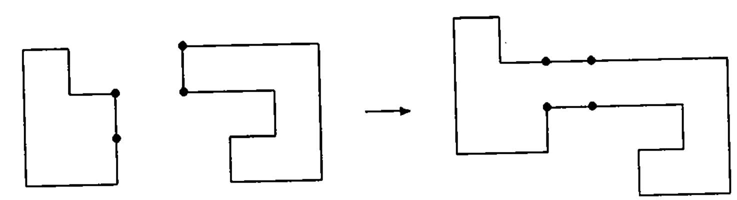


Figure 3.2: Concatenation of a 10-step polygon and a 14-step polygon to produce a 24-step polygon in $\mathbf{Z}^2$. The dots are the endpoints of the bonds that are changed during the concatenation.

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 $M, N \geq 4$,*

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

and

$$q_N \leq q_{N+2}. \tag{3.2.3}$$

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 $\mathbf{Z}^d$, as follows. We say that $(a_1, \ldots, a_d) \prec (b_1, \ldots, b_d)$ 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 $i = 1, \ldots, d$, let $e^{(i)}$ be the neighbour of the origin with $e_i^{(i)} = 1$ and $e_j^{(i)} = 0$ for $j \neq i$. For $i = 2, \ldots, d$ 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 $Q_2[M], \ldots, Q_d[M]$, and so, by symmetry,

$$|Q_2[M]| = \cdots = |Q_d[M]| \geq \frac{q_M}{d-1}. \tag{3.2.4}$$

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). □

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 $\lim_{n\to\infty}(q_{2n}/(d-1))^{1/2n}$ exists and equals some number $\mu_{Polygon} \leq \mu$, and that

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

for all even $N > 2$. In fact, $\mu_{Polygon} = \mu$; 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 $\mathbf{Z}^d$. There exists a constant K, depending only on the dimension d, such that for every integer $M \geq 1$,*

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

Proof. For a point x in $\mathbf{Z}^d$, 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 $\mathbf{R}^d$ which is orthogonal to the line containing 0 and x. Let i (respectively j) be chosen from among those values of $\{0, 1, \ldots, M\}$ that maximize (respectively, minimize) the dot product $\omega(i) \cdot \mathbf{v}$ (respectively, $v(j) \cdot \mathbf{v}$). Define

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

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

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

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

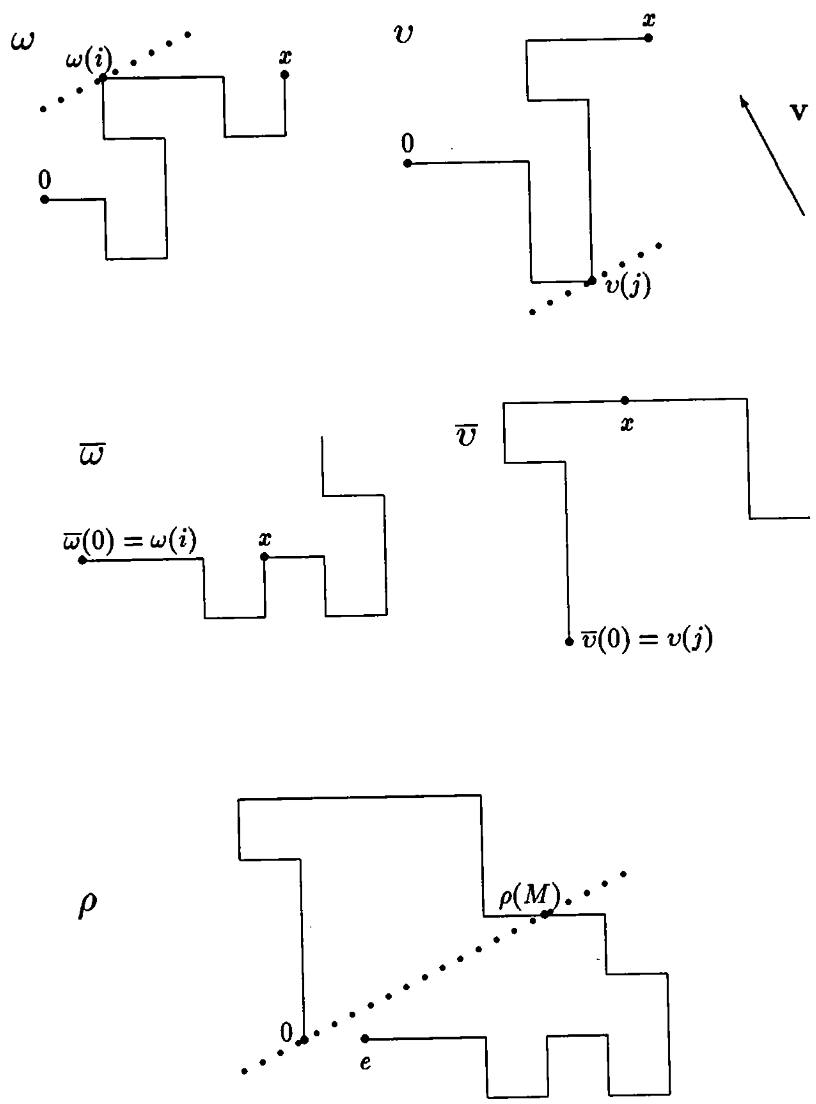


Figure 3.3: Proof of Theorem 3.2.4. Here $M = 12$. Top: the M-step bridges ω and υ, and the vector $\mathbf{v}$. Middle: the derived walks $\overline{\omega}$ and $\overline{\upsilon}$. Bottom: the $(2M+1)$-step walk ρ. The dotted lines are orthogonal to $\mathbf{v}$.

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 $\overline{\upsilon}$, followed by one step in the e direction, followed by the reversal of $\overline{\omega}$; that is,

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

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 $\rho(M) = x$.

Given a self-avoiding walk ϱ that has been constructed as above, we could reconstruct the original bridges ω and υ 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

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

The theorem is a direct consequence of (3.2.7). □

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

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

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

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

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 $\mu_{Polygon} \le \mu$. Finally, Equation (3.2.9) follows immediately from Equations (3.2.8) and (3.2.1). □

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

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 $N \to \infty$. To avoid trivialities, we also assume that $\|x^{(N)}\|_1$ has the same parity as N. Then*

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

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

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

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

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 $N \geq 2$*

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

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

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

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

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

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 $\{1, \ldots, n\}$ 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

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

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*

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

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

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

where

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

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

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

and the result follows. □

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 $i[0], \ldots, i[u]$, $k[0], \ldots, k[u] > 0$ (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 $\{1, \ldots, n - i[0]\}$ such that

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

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

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

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

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

and

$$i[j] < (n - i[0] - \cdots - i[j-1])^A, \tag{3.3.12}$$

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

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

Thus we end up with the inequality

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

Let $\mathcal{I}$ be the set of j's in $\{0, 1, \ldots, u-1\}$ 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.

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

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

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

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

$$h_n \leq n^{2|\mathcal{I}|+2}\mu^{i[0]+\cdots+i[u]} h^*_{n-i[0]-\cdots-i[u],k[u]}. \tag{3.3.17}$$

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

$$|\mathcal{I}| \leq Cn^{1-A}. \tag{3.3.18}$$

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

$$n2^{-a} \geq n - i[0] - \cdots - i[j-1] \geq n2^{-a-1}. \tag{3.3.19}$$

If we can show that

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

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

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

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),

$$n2^{-a} \geq n - i[0] - \cdots - i[f_a - 1] \geq n - i[0] - \cdots - i[F_a - 1] \geq n2^{-a-1}, \tag{3.3.22}$$

and hence

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

Also, (3.3.15) and (3.3.19) imply that

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

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 $m = n - i[0] - \cdots - i[u]$. (Recall that $k[u]$ and $i[u]$ are strictly positive.) Then (3.3.17) and (3.3.18) tell us that

$$h_n \leq \mu^{n-m} n^{2Cn^{1-A}+2} h^*_{m,k}. \tag{3.3.25}$$

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

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

and

$$l < (m+l)^A \leq n^A. \tag{3.3.27}$$

Combining (3.3.26) with Lemma 3.3.3 yields

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

which implies

$$h^*_{m,k} \leq \mu^{m+dV} n^2 [Dn]^{12n^B + 3dV}. \tag{3.3.29}$$

Applying (3.3.29) to (3.3.25), we obtain

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

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

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

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

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

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

$$J(Y) \leq 4m^B. \tag{3.3.33}$$

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

$$0 \leq g \leq dV. \tag{3.3.34}$$

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

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

(this is because (*i*) $\rho_1(i) = \eta_1(i) \in (0,k]$ whenever $0 < i \le m$, (*ii*) $\eta_1(m) \le \rho_1(i) \le Y_1$ whenever $m \le i \le m+g$, and (*iii*) $\rho_1(i) = Y_1 + \beta_1(i-m-g) \in [Y_1, Y_1+k]$ whenever $m+g \le i \le m+g+l$). 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 $(\rho(m+1), \ldots, \rho(m+g-1))$ with the rest of ρ; therefore

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

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 $r + a_1 + \cdots + a_t = |\rho|$.

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

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

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 $|\varphi| + a_1 + \cdots + a_{j-1}$ 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 $n \ge 2$ we have $q_n \le (d-1)\mu^n$ [by (3.2.5) and (3.2.9) for $n \ge 4$; the case $n = 2$ is obvious since $\mu \ge d$], and so the product (3.3.37) can be bounded above by

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

Therefore the number of possible walks ρ is at most

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

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 $b_r \le \mu^r$ [Equation (1.2.17)] that

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

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

3.4 Notes

Section 3.1. An earlier paper [Hammersley (1961b)] proved that $c_N \le \mu^N \exp[O(N^{(d-1)/d} \log N)]$ for $d \ge 2$. 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 $\mathbf{Z}^2$ 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 $\mu_{Polygon} = \mu$, 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 $\mathbf{Z}^2$. 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 $\mathbf{R}^2$ whose boundary is the simple closed curve determined by the polygon.) They conjecture that the analogous probability for any other point (a, b) of $\mathbf{R}^2$ (with a and b non-integer) should likewise increase to 1/2 as $N \to \infty$, 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:

$$m(z) = \liminf_{n\to\infty} \frac{-\log G_z(0, (n, 0, \ldots, 0))}{n}. \tag{4.1.1}$$

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 norm[1] $|\cdot|_z$. 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" $\mathsf{B}(z) \equiv \sum_x G_z(0, x)^2$ (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.

[1] Alternatively, one can work with the Euclidean norm and a direction-dependent mass $m[v; z] \equiv m(z)|v|_z$ for vectors $v \in R^d$ 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

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

with $\overline{\nu} = \nu$ (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 $(\omega(0), \ldots, \omega(N))$ whose steps $\omega(i) - \omega(i-1)$ all lie in a finite subset Ω of $\mathbf{Z}^d$ which is invariant under all symmetries of $\mathbf{Z}^d$. 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,

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

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

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 lim inf 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. □

Next, we shall show how to replace the "lim inf" 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 $\mathsf{B}(z) = \sum_x G_z(0, x)^2$. We will use the notation $(n, 0)$ to denote the point $(n, 0, \ldots, 0) \in \mathbf{Z}^d$; this notation will be generalized in Definition 4.1.7 below.

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

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

in particular, the limit exists and satisfies

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

(b) If $\mathsf{B}(z_c)$ 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 $\mathbf{Z}^d$,*

$$G_z(0, x) G_z(x, y) \leq \mathsf{B}(z) G_z(0, y). \tag{4.1.5}$$

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 we want.

Let (ω_A, ω_B) be a member of $\mathcal{S}_N$. 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 $\mathcal{T}_N$. This mapping is clearly one-to-one, so the lemma is proven. □

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

$$\lim_{n\to\infty} \frac{h_n(z)}{n} = \inf_{n\ge 1} \frac{h_n(z)}{n}, \tag{4.1.6}$$

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 $\mathsf{B}(z_c)$ is finite. Since $m(z)$ is nonincreasing, it suffices to show that

$$\limsup_{z\nearrow z_c} m(z) \le m(z_c). \tag{4.1.7}$$

Assume $\mathsf{B}(z_c)$ is finite. This implies that $G_{z_c}(0,x)$ is finite (for every x); hence, since $\mathsf{B}(z)$ 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 $m(z) \le h_n(z)/n$ for all z in $(0, z_c)$ [by (4.1.3)], this implies that

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

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

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 $\mathsf{B}(z_c) < \infty$ holds. However, it is expected that $\mathsf{B}(z_c)$ 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 $\mathbf{Z}^d$,*

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

Proof. Given x in $\mathbf{Z}^d$, 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 $j \ne i$, 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

$$G_z(0,x)G_z(x, x+u) \le \mathsf{B}(z)G_z(0, x+u) = \mathsf{B}(z)G_z(0,(2K,0)). \tag{4.1.10}$$

Applying symmetry and (4.1.4), we obtain $G_z(0,x)^2 \le \mathsf{B}(z)\exp[-m(z)2K]$, and (4.1.9) follows. □

Theorem 4.1.6 *If $\mathsf{B}(z_c) < \infty$, 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 $m(z_c) \ge 0$. 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. □

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 $y = (y_1, \ldots, y_{d-1})$ be a point of $\mathbf{Z}^{d-1}$, and let L be a nonnegative integer. Then (L, y) denotes the point $(L, y_1, \ldots, y_{d-1})$ in $\mathbf{Z}^d$, 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*

$$b_{N,L} = \sum_{y\in Z^{d-1}} b_{N,L}(y).$$

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

$$B_z(L) = \sum_{y \in Z^{d-1}} B_z(L, y).$$

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*

$$M(z) = \lim_{L\to\infty} \frac{-\log B_z(L,0)}{L} \text{ and } \overline{M}(z) = \lim_{L\to\infty} \frac{-\log B_z(L)}{L} \tag{4.1.11}$$

exist in $[-\infty, +\infty)$, and satisfy

$$B_z(L,0) \le e^{-LM(z)} \text{ and } B_z(L) \le e^{-L\overline{M}(z)} \tag{4.1.12}$$

for every integer $L \ge 1$. Also, $M(z)$ and $\overline{M}(z)$ are nonincreasing functions of z, and they satisfy the obvious inequalities $M(z) \ge \overline{M}(z)$ and $M(z) \ge m(z)$.

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

$$\sum_{n=0}^{N} b_{n,L_1} b_{N-n,L_2} \le b_{N,L_1+L_2}$$

and

$$\sum_{n=0}^{N} b_{n,L_1}(0) b_{N-n,L_2}(0) \le b_{N,L_1+L_2}(0)$$

for every nonnegative integer N, so

$$B_z(L_1)B_z(L_2) \le B_z(L_1 + L_2) \tag{4.1.13}$$

and

$$B_z(L_1,0)B_z(L_2,0) \le B_z(L_1 + L_2, 0) \tag{4.1.14}$$

for all $z > 0$. Therefore, by Lemma 1.2.2, the limits in (4.1.11) exist and satisfy (4.1.12) for all $L \ge 1$. □

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) $M(z) \le -\log z$ for all $z > 0$.*
(b) $M(z) \ge -\log(\mu z)$ 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 $B_z(L,0) \ge z^L$. The result follows now from (4.1.12).
(b) Fix z in $(0, z_c)$. Since $b_N \le \mu^N$ by (1.2.17),

$$B_z(L,0) \le \sum_{N=L}^{\infty} \mu^N z^N = \frac{(\mu z)^L}{1-\mu z}$$

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 $L \to \infty$. □

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 $\mathbf{Z}^{d-1}$, let $b^T_{N,L}(y)$ be the number of N-step bridges ω having $\omega(0) = 0$, $\omega(N) = (L, y)$, and $|\omega_i(k)| \le T$ for every $i = 2, \ldots, d$ and $k = 0, \ldots, N$. For real $z > 0$, let*

$$B^T_z(L,y) = \sum_N b^T_{N,L}(y) z^N.$$

Observe that the Monotone Convergence Theorem implies that for every $z > 0$

$$\lim_{T\to\infty} B^T_z(L,y) = B_z(L,y) \text{ for all } L \text{ and } y. \tag{4.1.15}$$

Also, by the usual concatenation argument, for every T we have

$$B^T_z(L_1,0)B^T_z(L_2,0) \le B^T_z(L_1+L_2,0) \text{ for all } L_1 \text{ and } L_2, \tag{4.1.16}$$

which implies, by Lemma 1.2.2, that we can define the "truncated masses"

$$M^T(z) = \lim_{L\to\infty} \frac{-\log B^T_z(L,0)}{L} = \inf_{L\ge 1} \frac{-\log B^T_z(L,0)}{L}. \tag{4.1.17}$$

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) $\lim_{T\to\infty} M^T(z) = \inf_{T\geq 1} M^T(z) = M(z)$.
(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 $L \to \infty$ [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. □

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 $\lim_{T\to\infty}\overline{M}^T(z) = \overline{M}(z)$. Observe that $\overline{B}_z^T(L,y) = 0$ if $||y||_\infty > LT$.

For any $L > 0$ and any y in $\mathbf{Z}^{d-1}$, 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 $T > 0$,

$$\begin{aligned} B_z(2L+1,0) &\geq \sum_{y\in Z^{d-1}:||y||_\infty\leq LT} z(\overline{B}_z^T(L,y))^2 \\ &\geq \frac{z(\sum_{y:||y||_\infty\leq LT}\overline{B}_z^T(L,y))^2}{\sum_{y:||y||_\infty\leq LT} 1^2} \\ &= \frac{z(\overline{B}_z^T(L))^2}{(2LT+1)^{d-1}}. \end{aligned}$$

This implies that for every fixed T, $M(z) \leq \overline{M}^T(z)$. The lemma follows. □

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. □

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

$$G_z(0,(n,0)) \leq (\chi(z))^2 B_z(n) \leq (\chi(z))^2 e^{-M(z)n}, \tag{4.1.18}$$

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. □

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. □

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. □

Corollary 4.1.17 *Define*

$$G_z(L) = \sum_{y \in Z^{d-1}} G_z(0,(L,y)), \tag{4.1.19}$$

the generating function of all self-avoiding walks from the origin to the hyperplane $x_1 = L$. Then

$$\lim_{L\to\infty} \frac{-\log G_z(L)}{L} = m(z) \text{ for all } z \text{ in } (0, z_c). \tag{4.1.20}$$

In fact,

$$e^{-m(z)L} \leq G_z(L) \leq \chi(z)^2 e^{-m(z)L} \text{ for all } L \geq 1. \tag{4.1.21}$$

Proof. Equation (4.1.20) follows from

$$B_z(L) \leq G_z(L) \leq \chi(z)^2 B_z(L), \tag{4.1.22}$$

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. □

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 $\mathsf{B}(z) = \sum_x G_z(0,x)^2$ is finite for $0 < z < z_c$ (since $\mathsf{B}(z) \leq \chi(z)^2$).

Theorem 4.1.18 *For any $0 < z < z_c$, there exists a norm $|\cdot|_z$ on $\mathbf{R}^d$, satisfying $\|u\|_\infty \leq |u|_z \leq \|u\|_1$ for every u in $\mathbf{R}^d$, such that*

$$\lim_{|x|_z\to\infty} \frac{-\log G_z(0,x)}{|x|_z} = m(z) \tag{4.1.23}$$

and

$$G_z(0,x) \leq \mathsf{B}(z)e^{-m(z)|x|_z} \quad \text{for every } x \text{ in } \mathbf{Z}^d. \tag{4.1.24}$$

Proof. Fix z in $(0, z_c)$. For each v in $\mathbf{Z}^d$, Lemma 4.1.4 tells us that

$$\frac{G_z(0,jv)}{\mathsf{B}(z)} \frac{G_z(0,kv)}{\mathsf{B}(z)} \leq \frac{G_z(0,(j+k)v)}{\mathsf{B}(z)}$$

for all nonnegative integers j and k. Therefore Lemma 1.2.2 implies that the limit

$$\lim_{n\to\infty} \frac{-\log G_z(0,nv)}{n}$$

exists, and, if we denote this limit by $m[v;z]$, that

$$G_z(0,nv) \leq \mathsf{B}(z)e^{-nm[v;z]} \quad \text{for every } n \geq 1. \tag{4.1.25}$$

We have $m[0;z] = 0$ since $G_z(0,0) = 1$, but for every nonzero v in $\mathbf{Z}^d$ we have $0 < m[v;z] \leq \|v\|_1 |\log z|$ (the lower bound follows from (1.3.14), while

$$G_z(0,v) \geq z^{\|v\|_1} \tag{4.1.26}$$

gives the upper bound). Also, it follows immediately that

$$m[kv;z] = |k| m[v;z] \tag{4.1.27}$$

for every integer k and every v in $\mathbf{Z}^d$, and (from Lemma 4.1.4) that

$$m[u+v;z] \le m[u;z] + m[v;z]$$

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 $|u+v|_z \le |u|_z + |v|_z$ for every u and v in $\mathbf{Q}^d$; in particular, then, we have $|v|_z \le ||v||_1$ for every v in $\mathbf{Q}^d$. Consequently, we obtain

$$|\,|u|_z - |v|_z| \le |u-v|_z \le ||u-v||_1;$$

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 $m[v;z] \ge m(z)||v||_\infty$ for every v in $\mathbf{Z}^d$. From this it follows that $||v||_\infty \le |v|_z$ on $\mathbf{Z}^d$, and hence on all of $\mathbf{R}^d$.

From (4.1.25) and the definition of $|\cdot|_z$, we obtain

$$G_z(0,v) \le \mathsf{B}(z)\exp(-|v|_z m(z)) \quad \text{for every } v \text{ in } \mathbf{Z}^d, \tag{4.1.28}$$

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

$$\lim_{n\to\infty} \frac{-\log G_z(0,x_n)}{|x_n|_z} > m(z). \tag{4.1.29}$$

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

$$||t - J^{-1}v||_1 \le \epsilon \quad \text{and} \quad |J^{-1}v|_z \le 1+\epsilon.$$

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

$$G_z(0,k(n)v)G_z(k(n)v,x_n) \le \mathsf{B}(z)G_z(0,x_n).$$

Now take logs of this inequality, divide by $-|x_n|_z$, and let n tend to infinity. Using

$$\begin{aligned}\lim_{n\to\infty} \frac{-\log G_z(0,k(n)v)}{|x_n|_z} &= J^{-1}m[v;z] \\ &= |J^{-1}v|_z m(z) \\ &\le (1+\epsilon)m(z)\end{aligned}$$

and [with the help of (4.1.26)]

$$\begin{aligned}\lim_{n\to\infty} \frac{-\log G_z(k(n)v,x_n)}{|x_n|_z} &\le -\lim_{n\to\infty} \frac{||k(n)v - x_n||_1}{|x_n|_z}\log z \\ &= -||J^{-1}v - t||_1 \log z \\ &\le -\epsilon \log z,\end{aligned}$$

we conclude that

$$\lim_{n\to\infty} \frac{-\log G_z(0,x_n)}{|x_n|_z} \le (1+\epsilon)m(z) - \epsilon\log z.$$

Since ϵ can be made arbitrarily close to 0, this contradicts our choice of $\{x_n\}$ and completes the proof the theorem. □

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 $G_z(0,(L,0,\ldots,0))$ 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 $0 < i \le N$ and

$$\omega_1(j) \le \omega_1(i) < \omega_1(k) \text{ for all } j = 0,\ldots,i \text{ and } k = i+1,\ldots,N. \tag{4.2.1}$$

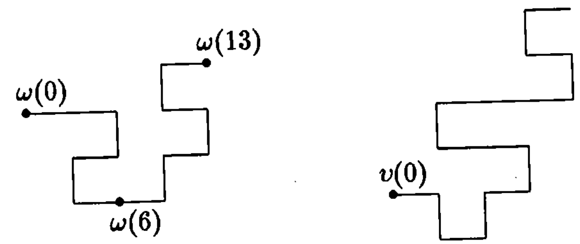


Figure 4.1: The bridge ω on the left can be decomposed into two smaller bridges: a 6-step bridge followed by a 7-step bridge. The bridge v on the right is irreducible.

Then ω can be decomposed into two smaller bridges, $(\omega(0),\ldots,\omega(i))$ and $(\omega(i),\ldots,\omega(N))$. (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 $a = 1,\ldots,\omega_1(N)$, 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 $(\omega(0),\ldots,\omega(s))$ is an irreducible bridge and $(\omega(s),\ldots,\omega(N))$ is a bridge. It is thus straightforward to see that

$$b_N = \sum_{s=1}^{N} \lambda_s b_{N-s} + \delta_{N,0} \tag{4.2.2}$$

for every $N \geq 0$. From this equation we immediately obtain

$$B_z = \frac{1}{1-\Lambda_z} \text{ for all complex } z \text{ with } |z| < z_c, \tag{4.2.3}$$

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,

$$\sum_{k=1}^{\infty} \frac{\lambda_k}{\mu^k} = 1. \tag{4.2.4}$$

(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 $N \geq 1$,

$$a_N = \sum_{k=1}^{N} p_k a_{N-k}. \tag{4.2.5}$$

To interpret this probabilistically, suppose that we have an independent sequence of random variables $X_1, X_2, \ldots$ with common distribution $\Pr\{X = k\} = p_k$. Then

$$a_N = \Pr\{X_1 + \ldots + X_k = N \text{ for some } k \geq 0\}; \tag{4.2.6}$$

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 $v_0, v_1, \ldots$ by

$$\begin{aligned} v_0 &= g_0 \\ v_n &= g_n + f_1 v_{n-1} + f_2 v_{n-2} + \cdots + f_n v_0, \text{ for all } n \geq 1. \end{aligned}$$

(a) If $f < 1$, then $\lim_{n\to\infty} v_n = 0$ and $\sum_{n=0}^{\infty} v_n = g/(1-f)$.
(b) If $f = 1$, then

$$\lim_{n\to\infty} v_n = \frac{g}{\sum_{k=1}^{\infty} k f_k}$$

(the limit is 0 if the sum in the denominator diverges). Also, $\sum_n v_n$ diverges.
(c) If $f > 1$, then $\limsup_{n\to\infty} v_n^{1/n} > 1$.

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 $\lim_{N\to\infty} b_N/\mu^N$ exists and equals $(\sum_{k=1}^{\infty} kp_k)^{-1}$. 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 $\mathbf{Z}^{d-1}$ 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*

$$\lambda_{N,L} = \sum_{y\in Z^{d-1}} \lambda_{N,L}(y).$$

For each real $z > 0$, we define the generating functions

$$\Lambda_z(L,y) = \sum_{N=1}^{\infty} \lambda_{N,L}(y) z^N \quad \textit{and} \quad \Lambda_z(L) = \sum_{N=1}^{\infty} \lambda_{N,L} z^N.$$

Using this definition, we have the following refinements of (4.2.2):

$$b_{N,L} = \sum_{k=1}^{N}\sum_{j=1}^{L} \lambda_{k,j} b_{N-k,L-j} + \delta_{N,0}\delta_{L,0} \tag{4.2.7}$$

$$b_{N,L}(y) = \sum_{k=1}^{N}\sum_{j=1}^{L}\sum_{v\in Z^{d-1}} \lambda_{k,j}(v) b_{N-k,L-j}(y-v) + \delta_{N,0}\delta_{L,0}\delta_{y,0}. \tag{4.2.8}$$

These imply the following equations for the generating functions:

$$B_z(L) = \sum_{j=1}^{L} \Lambda_z(j) B_z(L-j) + \delta_{L,0} \tag{4.2.9}$$

$$B_z(L,y) = \sum_{j=1}^{L}\sum_{v\in Z^{d-1}} \Lambda_z(j,v) B_z(L-j, y-v) + \delta_{L,0}\delta_{y,0}. \tag{4.2.10}$$

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

$$p_L \equiv p_L(z) = \Lambda_z(L) e^{m(z)L} \qquad (L \ge 1) \tag{4.2.11}$$

and

$$a_L \equiv a_L(z) = B_z(L) e^{m(z)L} \qquad (L \ge 0). \tag{4.2.12}$$

Observe that $p_L \le a_L \le 1$ [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

$$P(s) = \sum_{k=1}^{\infty} p_k s^k \quad \text{and} \quad A(s) = \sum_{n=0}^{\infty} a_n s^n; \tag{4.2.13}$$

these are finite for $|s| < 1$. From (4.2.9) we then get

$$A(s) = \frac{1}{1-P(s)} \text{ for } |s| < 1. \tag{4.2.14}$$

The sequence a_L is bounded away from zero [in fact, $a_L \ge (\chi(z))^{-2}$ 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,

$$\sum_{k=1}^{\infty} p_k = 1. \tag{4.2.15}$$

In addition, since we have just seen that a_L is bounded away from 0, the renewal theorem tells us that

$$\lim_{L\to\infty} a_L = \frac{1}{\sum_{k=1}^{\infty} kp_k} > 0. \tag{4.2.16}$$

Therefore

$$\sum_{k=1}^{\infty} kp_k < +\infty. \tag{4.2.17}$$

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

$$\limsup_{k\to\infty} p_k^{1/k} < 1. \tag{4.2.18}$$

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*

$$m_\Lambda(z) = \liminf_{L\to\infty} \frac{-\log \Lambda_z(L)}{L}. \tag{4.2.19}$$

Then

$$m_\Lambda(z) > m(z). \tag{4.2.20}$$

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

$$\Lambda_z(L) = \sum_{N=3L}^{\infty} \lambda_{N,L} z^N \leq \sum_{N=3L}^{\infty} \mu^N z^N = \frac{(\mu z)^{3L}}{1-\mu z}, \tag{4.2.21}$$

so

$$m_\Lambda(z) \geq -3\log(\mu z). \tag{4.2.22}$$

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

$$m_\Lambda(z) \leq -3\log z \tag{4.2.23}$$

(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*

$$\left| B_z(L)e^{m(z)L} - C_z \right| \leq e^{-\epsilon(z)L}$$

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 $\mathrm{Res}(f(s), a)$. By the residue theorem and the integral theorem for the coefficients of a Laurent series, we have

$$a_L = -\mathrm{Res}\left(\frac{s^{-L-1}}{1-P(s)}, 1\right) + \frac{1}{2\pi i}\oint_{|s|=r} \frac{s^{-L-1}}{1-P(s)} ds.$$

The first term on the right is $1/P'(1)$ and the second is $O(r^{-L-1})$. This proves the theorem. □

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*

$$\lim_{L\to\infty} \left| B_z(L,y)e^{m(z)L} L^{(d-1)/2} - C_z \frac{\exp(-|y|^2/\delta_z L)}{(\pi\delta_z)^{(d-1)/2}} \right| = 0 \qquad (4.2.24)$$

uniformly in y in $\mathbf{Z}^{d-1}$. Consequently,

$$B_z(L,y) \sim C_z e^{-m(z)L} \frac{\exp(-|y|^2/\delta_z L)}{(\pi\delta_z L)^{(d-1)/2}} \qquad (4.2.25)$$

as $L \to \infty$, uniformly in y in $\mathbf{Z}^{d-1}$ satisfying $|y| \le KL^{1/2}$ (for every fixed $K > 0$). Equivalently we have (by Theorem 4.2.5)

$$\frac{B_z(L,y)}{B_z(L)} \sim \frac{\exp(-|y|^2/\delta_z L)}{(\pi\delta_z L)^{(d-1)/2}} \qquad (4.2.26)$$

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 $\{1,2,\ldots\}$ and Y takes values in $\mathbf{Z}^{d-1}$, and

$$\Pr\{(X,Y) = (L,y)\} = \Lambda_z(L,y)e^{m(z)L} \text{ for } L \ge 1, y \in \mathbf{Z}^{d-1}. \qquad (4.2.27)$$

Thus the marginal distribution of X is given by (4.2.11):

$$\Pr\{X = L\} = p_L(z), \text{ for } L = 1,2,\ldots$$

(In particular, (4.2.15) guarantees that (4.2.27) describes a genuine probability distribution.) Let $\{(X_n,Y_n) : n \ge 1\}$ be a sequence of independent copies of (X,Y). Then we claim that

$$B_z(L,y)e^{m(z)L} = \Pr\{X_1 + \cdots + X_k = L \text{ and } Y_1 + \cdots + Y_k = y \text{ for some } k\} \qquad (4.2.28)$$

for $L \ge 1$ 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

$$B_z(L,y) = \sum_{k=1}^{L} \left[\sum \prod_{i=1}^{k} \Lambda_z(n_i, v_i) \right]$$

where the inner sum is over all $n_1,\ldots,n_k \ge 1$ and all $v_1,\ldots,v_k$ in $\mathbf{Z}^{d-1}$ such that $n_1 + \ldots + n_k = L$ and $v_1 + \ldots + v_k = y$.

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 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 $L \ge 1$. 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

$$\begin{aligned} E(e^{s|Y|}) &= \sum_{L,y} e^{s|y|}\Lambda_z(L,y)e^{m(z)L} \\ &\le \sum_{L,y,N} e^{sN}\lambda_{N,L}(y)z^N e^{m(z)L} \\ &= \sum_L \Lambda_{ze^s}(L)e^{m(z)L} < \infty. \end{aligned}$$

□

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 $\{(X_n,Y_n) : n \ge 1\}$ 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:

$$(X_n, Y_n) = (X_n, Y_{n2}, Y_{n3}, \ldots, Y_{nd}).$$

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 $i = 2, \ldots, d$ 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

$$(X_1 + \cdots + X_n, Y_1 + \cdots + Y_n) = (L, y)$$

[i.e. the expected number of times that this d-dimensional random walk visits the point (L, y)]. Then Stam's theorem says that

$$\lim_{L \to \infty} |L^{(d-1)/2} U(L, y) - \varphi(L, y)| = 0,$$

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 $L \to \infty$, 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 $\delta_z = 2v/\theta$. □

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

$$f(u, z) \equiv \sum_{L=1}^{\infty} \sum_{N=1}^{\infty} \lambda_{N,L} z^N u^L = \sum_{L=1}^{\infty} \Lambda_z(L) u^L. \qquad (4.2.29)$$

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

$$\sum_{L=1}^{\infty} \sum_{N=1}^{\infty} \lambda_{N,L} |z|^N |u|^L \leq \sum_{L=1}^{\infty} \Lambda_{z_0+\epsilon}(L) (e^{m(z_0)} + \epsilon)^L. \qquad (4.2.30)$$

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

$$e^{-m_\Lambda(z_0+\epsilon)} \left(e^{m(z_0)} + \epsilon\right) < 1. \qquad (4.2.31)$$

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

$$f(e^{m(z)}, z) = 1 \quad \text{for all } z \text{ in } (0, z_c); \qquad (4.2.32)$$

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. □

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* $\omega = (\omega(0), \ldots, \omega(N))$ *be a bridge. A* backtrack *of* ω *is a subwalk*

$$\omega[s;t] \equiv (\omega(s), \ldots, \omega(t))$$

$(0 \le s < t \le N)$ *satisfying:*
(i) $\omega_1(t) \le \omega_1(i) < \omega_1(s)$ *for all* $i = s+1, \ldots, t-1$;
(ii) $\omega_1(j) \le \omega_1(s)$ *for all* $j = 0, 1, \ldots, s-1$; *and*
(iii) $\omega_1(t) < \omega_1(j)$ *for all* $j = t+1, \ldots, N$.
The span *of the backtrack is defined to be* $\omega_1(s) - \omega_1(t)$.

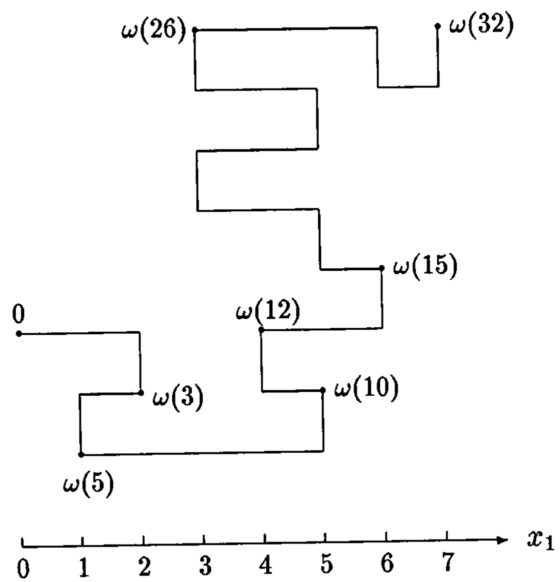


Figure 4.2: A bridge of length $N = 32$, with $\omega_1(0) = 0$ and $\omega_1(N) = 7$. There are three backtracks: $\omega[3;5]$, $\omega[10;12]$, and $\omega[15;26]$.

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:

Lemma 4.3.2 *Let* ω *be an* N*-step bridge. If* $0 \le s_0 < t_0 \le N$ *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

$$\begin{aligned} A &= \min\{\omega_1(i) : t_0 \le i \le N\}, \\ B &= \max\{\omega_1(i) : 0 \le i \le s_0\}, \\ s &= \max\{i : 0 \le i \le s_0, \omega_1 = B\}, \\ t &= \max\{i : t_0 \le i \le N, \omega_1 = A\}. \end{aligned}$$

Then $\omega[s;t]$ is the unique backtrack containing $\omega(s_0)$ and $\omega(t_0)$. Details are left to the reader. □

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* $\omega_1(t) \le j < \omega_1(s)$. *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* $\{1, \ldots, N-1\}$ *such that* $\omega_1(i) \le j$ *for all* $i = 0, \ldots, r$ *and* $\omega_1(i) > j$ *for all* $i = r+1, \ldots, N$.

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 $c+1,\ldots,c+T-1$. Then there exists a decreasing function $\epsilon(T)$ (independent of L and c) such that $\lim_{T\to\infty}\epsilon(T) = 0$ and*

$$B^*(L;c,T) \leq e^{-mL}\epsilon(T) \text{ for all } L \text{ and } c \ (0 \leq c \leq L-T). \tag{4.3.1}$$

[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). □

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

$$A = \{Q, 2Q, \ldots, kQ\}.$$

With 0 and L, the elements of A are the endpoints of the blocks. Consider any nonempty subset $S = \{n_1,\ldots,n_\tau\}$ of A, with $n_1 < \ldots < n_\tau$. Put $n_0 = 0$ and $n_{\tau+1} = L$. Then

$$n_{i+1} - n_i \geq Q \text{ for all } i = 0,1,\ldots,\tau.$$

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 $\sigma_1,\ldots,\sigma_\tau \geq 1$, let $\mathcal{J}_L(S,\sigma_1,\ldots,\sigma_\tau)$ denote the set of all irreducible bridges of span L such that for each $i = 1,\ldots,\tau$ 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*

$$\bigcup_{S\subset A,|S|\geq k/2} \mathcal{I}_L(\leq\Delta;S) \tag{4.3.2}$$

and

$$\bigcup_{S\subset A,|S|\geq 1}\ \bigcup_{\substack{\sigma_1>\Delta,\ldots,\sigma_\tau>\Delta\\ \sigma_1+\cdots+\sigma_\tau>k\Delta/2}} \mathcal{J}_L(S,\sigma_1,\ldots,\sigma_\tau). \qquad (4.3.3)$$

Lemma 4.3.7 $GF(\mathcal{I}_L(\leq\Delta;S))\leq e^{-mL}(\chi\epsilon(T))^{\tau+1}$.

Here ϵ is from Lemma 4.3.5 and χ is the susceptibility [defined in (1.3.4)].

Lemma 4.3.8 $GF(\mathcal{J}_L(S,\sigma_1,\ldots,\sigma_\tau))\leq e^{-mL}\chi^{2\tau}e^{-m(\sigma_1+\cdots+\sigma_\tau)}$.

Proof of Theorem 4.2.4. Fix T so that

$$2\,(\chi\epsilon(T))^{1/2}<\frac{1}{2}, \qquad (4.3.4)$$

and fix Δ so that

$$\left(1+\frac{\chi^2}{1-e^{-m/2}}\right)e^{-m\Delta/4}<\frac{1}{2}. \qquad (4.3.5)$$

Set $Q=2\Delta+T$. Then, by the preceding three lemmas,

$$\begin{aligned}\Lambda_z(L) &\equiv GF(\mathcal{I}_L)\\ &\leq 2^k e^{-mL}\,[\chi\epsilon(T)]^{1+k/2}\\ &\quad + e^{-mL}\sum_{\tau=1}^{k}\binom{k}{\tau}\chi^{2\tau}\sum_{\substack{\sigma_1>\Delta,\ldots,\sigma_\tau>\Delta\\ \sigma_1+\cdots+\sigma_\tau>k\Delta/2}} e^{-m(\sigma_1+\cdots+\sigma_\tau)}.\end{aligned}$$

Now, for any positive D and r such that $r<m$,

$$\begin{aligned}&\sum_{\substack{\sigma_1>\Delta,\ldots,\sigma_\tau>\Delta\\ \sigma_1+\cdots+\sigma_\tau>D}} e^{-m(\sigma_1+\cdots+\sigma_\tau)}\\ &\leq \sum_{\sigma_1>\Delta,\ldots,\sigma_\tau>\Delta} e^{-m(\sigma_1+\cdots+\sigma_\tau)}e^{r(\sigma_1+\cdots+\sigma_\tau-D)}\\ &\leq \left(\frac{e^{(r-m)\Delta}}{1-e^{-(m-r)}}\right)^{\tau}e^{-rD}.\end{aligned}$$

Putting $r=m/2$ and $D=k\Delta/2$, we obtain [using (4.3.4) and (4.3.5)]

$$\begin{aligned}\Lambda_z(L) &\leq e^{-mL}\left(\left[2\,(\chi\epsilon(T))^{1/2}\right]^k+\left[1+\frac{\chi^2e^{-m\Delta/2}}{1-e^{-m/2}}\right]^k e^{-m\Delta k/4}\right)\\ &\leq e^{-mL}\left(\left(\frac{1}{2}\right)^k+\left(\frac{1}{2}\right)^k\right)\\ &\leq 2e^{-mL}2^{2-(L/Q)}.\end{aligned}$$

Therefore

$$m_\Lambda=\liminf_{L\to\infty}\frac{-\log\Lambda_z(L)}{L}\geq m+\frac{\log 2}{Q}>m.$$

This proves the theorem. □

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 $S=\{n_1,\ldots,n_\tau\}\subset A$ and some $\sigma_1,\ldots,\sigma_\tau>\Delta$ satisfying $\sigma_1+\cdots+\sigma_\tau>(k/2)\Delta$, ω is in $\mathcal{J}_L(S,\sigma_1,\ldots,\sigma_\tau)$. This proves the lemma. □

Proof of Lemma 4.3.7. Suppose that ω is in $\mathcal{I}_L(\leq\Delta;S)$, $S=\{n_1,\ldots,n_\tau\}$. Let

$$\begin{aligned}r_0 &= 0,\\ q_{\tau+1} &= |\omega|,\\ q_i &= \min\{j:\omega_1(j+1)=n_i+1\}, \quad i=1,\ldots,\tau,\\ r_i &= \max\{j:\omega_1(j)=n_i\}, \quad i=1,\ldots,\tau.\end{aligned}$$

(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

$$\omega_1(j)\geq n_i-\Delta \qquad \text{for all } j\geq q_i$$

and

$$\omega_1(j)\leq n_i+\Delta \qquad \text{for all } j\leq r_i.$$

In particular, $r_{i-1}<q_i$ $(i=1,\ldots,\tau+1)$, 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

$$\{x\in\mathbf{R}^d:n_{i-1}+\Delta<x_1<n_i-\Delta\}.$$

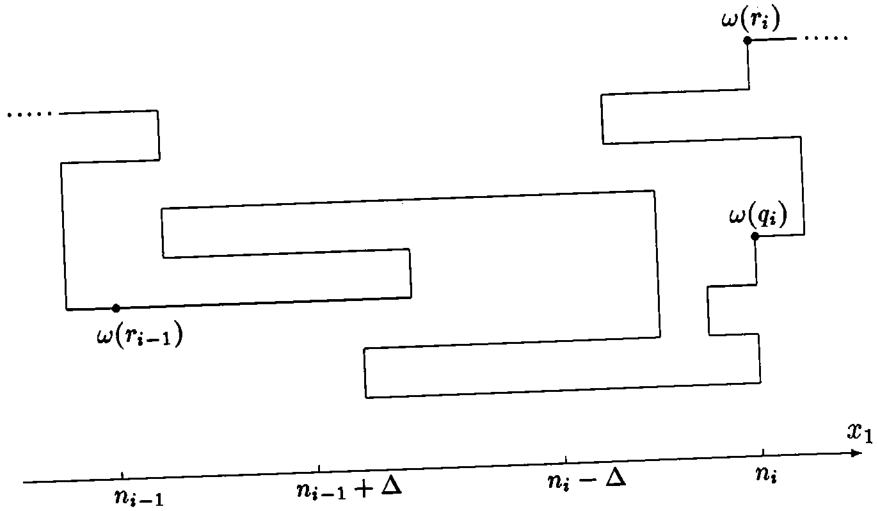


Figure 4.3: Proof of Lemma 4.3.7: part of a walk ω in $\mathcal{I}_L(\leq \Delta; S)$.

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 $q_1, r_1, \ldots, q_\tau, r_\tau$, 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. □

Proof of Lemma 4.3.8. Suppose ω is in $\mathcal{J}_L(S, \sigma_1, \ldots, \sigma_\tau)$. For every $i = 1, \ldots, \tau$, 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 $i = 1, \ldots, \tau - 1$.

Define $l_0 = 0$ and, for $i = 1, \ldots, \tau$,

$$\begin{aligned} f_i &= \min\{r > l_{i-1} : \omega_1(r+1) = n_i + 1\}, \\ l_i &= \max\{r : \omega_1(r) = n_i\}. \end{aligned}$$

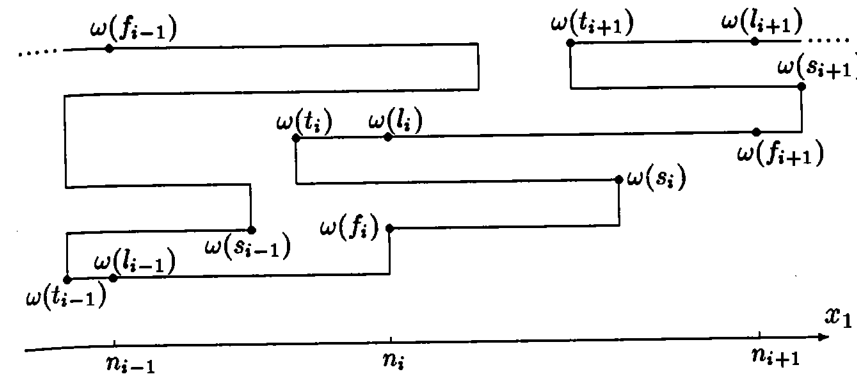


Figure 4.4: Proof of Lemma 4.3.8: part of a walk ω in $\mathcal{J}_L(S, \sigma_1, \ldots, \sigma_\tau)$.

Then, for every $i = 1, \ldots, \tau$,

$$l_{i-1} < f_i < s_i < t_i \leq l_i,$$

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

$$\begin{aligned} GF(\mathcal{J}_L(S, \sigma_1, \ldots, \sigma_\tau)) &\leq \left(\prod_{i=1}^{\tau+1} B(n_i - n_{i-1})\right) \prod_{i=1}^{\tau} (\chi B(\sigma_i)\chi) \\ &\leq e^{-mL}\chi^{2\tau} e^{-m(\sigma_1 + \cdots + \sigma_\tau)}, \end{aligned}$$

since $\sum_{i=1}^{\tau+1}(n_i - n_{i-1}) = L$. □

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 $\hat{\mathcal{B}}_{N,L,r}$ be the set of bridges $(\omega(0),\ldots,\omega(N))$ of span L such that $\omega(0)=0$, $\omega_2(N)=r$, and $0\le\omega_2(i)\le r$ for every $i=0,\ldots,N$. Also let $\hat{\mathcal{B}}_{N,L}=\cup_{r\ge 0}\hat{\mathcal{B}}_{N,L,r}$.*

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 $\hat{\mathcal{B}}_{N,L}$.

Proposition 4.4.2 *There exists an N_0 such that $b_{N,L}\le e^{3N^{1/2}}|\hat{\mathcal{B}}_{N,L}|$ for every $N\ge N_0$ and every $L\ge 1$.*

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

$$A_j=\max_{n_0<i\le N}(-1)^j(\omega_2(n_{j-1})-\omega_2(i))$$

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 $\overline{A}_j(\omega)$ and $m_j(\omega)$ so that

$$\overline{A}_j=\max_{0\le i<m_{j-1}}(-1)^j(\omega_2(m_{j-1})-\omega_2(i))$$

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 $\omega(m_j)$, $j=l-2,\ldots,0$, and at $\omega(n_j)$, $j=1,\ldots,k-1$. 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 $\hat{\mathcal{B}}_{N,L,A+\overline{A}}$, where $A=\sum A_j$ and $\overline{A}=\sum\overline{A}_j$. Since $A_1>A_2>$

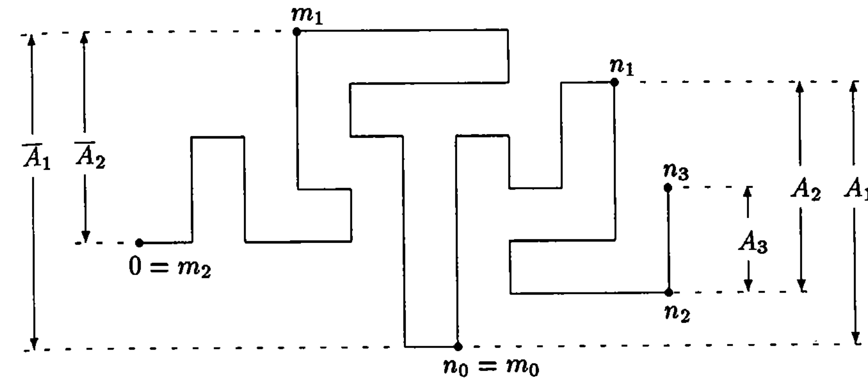


Figure 4.5: Proof of Proposition 4.4.2: The walk ω before "unfolding".

$\ldots>A_k$, there are $P_D(A)$ possible sequences of A_j's that sum to A; and since $\overline{A}_1\ge\overline{A}_2>\ldots>\overline{A}_l$, there are $P_D(\overline{A}+1)$ possible sequences of $\overline{A}_j$'s that sum to $\overline{A}$. Observe that these two sequences together with the final bridge in $\hat{\mathcal{B}}_{N,L,A+\overline{A}}$ determine the original ω uniquely, except perhaps for a reflection through $x_2=0$. Therefore, since $A+\overline{A}\le N$, we obtain

$$\begin{aligned}b_{N,L}&\le 2\sum_{r=0}^{N}\sum_{A=0}^{r}P_D(A)P_D(r-A+1)|\hat{\mathcal{B}}_{N,L,r}|\\&\le 2\sum_{r=0}^{N}(N+1)K^2\exp\left[(3-\epsilon)(N+1)^{1/2}\right]|\hat{\mathcal{B}}_{N,L,r}|,\end{aligned}$$

where the second inequality follows as in (3.1.7), having taken $B=3$ in (3.1.6). The proposition follows. □

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 $\hat{B}_z(L)=\sum_N|\hat{\mathcal{B}}_{N,L}|z^N$. Then*

$$\lim_{L\to\infty}\frac{-\log\hat{B}_z(L)}{L}=M(z)$$

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

$$\limsup_{L\to\infty} \frac{-\log \hat{B}_z(L)}{L} \leq M((1-\epsilon)z). \tag{4.4.1}$$

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 $\hat{B}_z(L) \leq B_z(L)$. □

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 $\cup_N \hat{\mathcal{B}}_{N,L}$. Let $\overline{\omega}_B$ 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 $\overline{\omega}_B$, 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). □

Corollary 4.4.5 $\lim_{N\to\infty}(\lambda_N)^{1/N} = \mu$.

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 $\hat{\mathcal{B}}_{n-1,L}$; then the resulting irreducible bridge ω has $3n$ steps. Thus $\lambda_{3n,L+1}$ is bounded below by $|\hat{\mathcal{B}}_{n-1,L}|^3$. 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

$$\lambda_{N,L+1} \geq |\hat{\mathcal{B}}_{\lfloor N/3\rfloor-1,L}|^3$$

(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 $L = 1, \ldots, N$ 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. □

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 that the integer j is a break point of ω if there exists an r in $\{1, \ldots, N-1\}$ such that $\omega_1(i) \leq j$ for all $i = 0, \ldots, r$ and $\omega_1(i) > j$ for all $i = r+1, \ldots, N$. 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 $(\omega(0), \ldots, \omega(N))$ such that: $\omega(0) = 0$; $\omega_1(N) = L$; $\omega_1(i) \leq L$ for every $i = 0, \ldots, N$; and ω has no break points. Then*

$$\liminf_{L\to\infty} \frac{-\log H_z(L)}{L} = m_\Lambda(z).$$

Proof. Since $H_z(L) \geq \Lambda_z(L)$ we have

$$\liminf_{L\to\infty} \frac{-\log H_z(L)}{L} \leq m_\Lambda(z),$$

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 $(\omega(T), \ldots, \omega(N))$ is a bridge of span L, and that if it has any break points, then they must all be in $\{1, 2, \ldots, J-1\}$. We cut ω at I, T, and the last break point of $(\omega(T), \ldots, \omega(N))$; 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

$$\begin{aligned} H_z(L) &\leq \sum_{J=0}^{L} [\chi(z)^2 e^{-m(z)J}]^2 \sum_{k=0}^{J-1} e^{-m(z)k} Re^{-\rho(L-k)} \\ &\leq \chi(z)^4 \sum_{J=0}^{L} e^{-2m(z)J} J Re^{(\rho - m(z))J} e^{-\rho L} \\ &\leq \chi(z)^4 RL^2 e^{-\rho L} \end{aligned}$$

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 $m_\Lambda(z)$. □

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 $\bar{\epsilon}(z)$ and $\overline{C}_z$ such that

$$\left| G_z(L) e^{m(z)L} - \overline{C}_z \right| \leq e^{-\bar{\epsilon}(z)L}$$

for all sufficiently large L.
(b) (Ornstein-Zernike decay) For δ_z as in Theorem 4.2.6,

$$\lim_{L\to\infty} \left| G_z(0, (L, y)) e^{m(z)L} L^{(d-1)/2} - \overline{C}_z \frac{\exp(-|y|^2/\delta_z L)}{(\pi \delta_z)^{(d-1)/2}} \right| = 0$$

uniformly in y in $\mathbf{Z}^{d-1}$.

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 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 $2G_z(\lfloor L/3 \rfloor)^2 G_z(L)$, 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

$$G_z(L) = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1-i} H_z(i) B_z(L-1-i-j) z H_z(j) + G_z^*(L) \qquad (4.4.2)$$

(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 $\overline{C}_z = ze^m C_z (\sum_{n\geq 0} H_z(n) e^{nm})^2$, 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

$G_z(L)e^{Lm} - \overline{C}_z =$

$$\sum_{i=0}^{L-1}\sum_{j=0}^{L-1-i} zH_z(i)[B_z(L-1-i-j)e^{(L-1-i-j)m} - C_z]H_z(j)e^{(i+j+1)m}$$
$$-C_z \sum_{i+j\geq L} zH_z(i)H_z(j)e^{(i+j+1)m} \; + \; G_z^*(L)e^{Lm}.$$

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 $\mathbf{Z}^d$, 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

$$G_z(0,(L,y)) =$$
$$\sum_{i=0}^{L-1}\sum_{j=0}^{L-1-i}\sum_{u,v\in Z^{d-1}} H_z(i,u)B_z(L-1-i-j, y-u-v)zH_z(j,v)$$
$$+ G_z^*(0,(L,y)). \tag{4.4.3}$$

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

$$\sup_{y\in Z^{d-1}} \Big|\sum_{k=1}^{L}\sum_{i=0}^{k-1}\sum_{u,v} zH_z(i,u)B_z(L-k, y-u-v)H_z(k-1-i,v)e^{Lm}L^{(d-1)/2}$$
$$- ze^m(\sum_{n=0}^{\infty} H_z(n)e^{nm})^2\varphi(L,y)\Big| \tag{4.4.4}$$

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

$$B_z(L,y)e^{Lm}L^{(d-1)/2} - \varphi(L,y) \to 0$$

uniformly in y as $L \to \infty$. 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

$$\sum_{k=1}^{L/2}\sum_{i=0}^{k-1}\sum_{u,v} zH_z(i,u)H_z(k-1-i,v)e^{km}(q_{L,k} + r_{L-k} + s_{L,k,u+v})$$

$$+ \sum_{k=L/2+1}^{L}\sum_{i=0}^{k-1}\sum_{u,v} zH_z(i,u)H_z(k-1-i,v)e^{km}L^{(d-1)/2}$$
$$+ \sum_{k=L/2+1}^{\infty}\sum_{i=0}^{k-1} zH_z(i)H_z(k-i-1)e^{km}\varphi(L,0),$$

where we define

$$q_{L,k} = \sup_{x\in Z^{d-1}} |L^{(d-1)/2} - (L-k)^{(d-1)/2}|B(L-k,x)e^{(L-k)m},$$

$$r_n = \sup_{x\in Z^{d-1}} |n^{(d-1)/2}B_z(n,x)e^{nm} - \varphi(n,x)|,$$

and

$$s_{L,k,w} = \sup_{y\in Z^{d-1}} |\varphi(L-k, y-w) - \varphi(L,y)|.$$

By Lemma 4.4.6, there exist positive constants A and ϵ (depending only on z) such that

$$\sum_{i=0}^{k-1}\sum_{u,v} H_z(i,u)H_z(k-1-i,v)e^{km} = \sum_{i=0}^{k-1} H_z(i)H_z(k-1-i)e^{km} \leq Ae^{-\epsilon k} \tag{4.4.5}$$

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 $L \to \infty$ 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, $\lim_{n\to\infty} r_n = 0$ 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

$$q_{L,k} \leq \left|\left(\frac{L}{L-k}\right)^{(d-1)/2} - 1\right|\Psi. \tag{4.4.6}$$

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

$$\begin{aligned} |\exp(-|y|^2/\delta L) - \exp(-|y-w|^2/\delta L)| &\leq \sup_{t\in R}\left|\frac{d}{dt}e^{-t^2/\delta}\right|\left|\frac{|y|}{\sqrt{L}} - \frac{|y-w|}{\sqrt{L}}\right| \\ &\leq \sup_{t\in R}\left|\frac{2t}{\delta}e^{-t^2/\delta}\right|\frac{|w|}{\sqrt{L}} \\ &\leq \text{const.}\frac{|w|}{\sqrt{L}}. \end{aligned}$$

Secondly, for arbitrary $\zeta \geq 0$ (so in particular for $\zeta = |y-w|^2/\delta$):

$$\begin{aligned} |\exp(-\zeta/L) - \exp(-\zeta/(L-k))| &= \exp(-\zeta/L)|1-\exp(-\zeta k/L(L-k))| \\ &\leq \exp(-\zeta/L)\frac{\zeta k}{L(L-k)} \\ &\leq \left(\sup_{t\geq 0} te^{-t}\right)\frac{k}{L-k} \end{aligned}$$

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 $y \in \mathbf{Z}^{d-1}$ [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.