

Exact Computations in the Statistical Mechanics of Disordered Systems

by
LAWRENCE K. SAUL

A. B. in Physics
Harvard College
(June 1990)

SUBMITTED TO THE DEPARTMENT OF
PHYSICS
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE
DEGREE OF
DOCTOR OF PHILOSOPHY IN PHYSICS

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 1994

© 1994 Massachusetts Institute of Technology

Signature of Author:

Department of Physics
May 9, 1994

Certified by:

Mehran Kardar
Professor of Physics
Thesis Supervisor

Accepted by:

George F. Koster, Chairman
Departmental Graduate Committee
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

MAY 25 1994

Exact Computations in the Statistical Mechanics of Disordered Systems

by

LAWRENCE K. SAUL

Submitted to the Department of Physics
on May 9, 1994 in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy in Physics.

ABSTRACT

We investigate the potential for exact computations in the statistical mechanics of disordered systems. Three systems are examined: directed waves in random media, the 2D $\pm J$ Ising spin glass, and tree-like neural networks.

Unitary propagation of directed waves is described by a Schrödinger equation with a random time-dependent potential. We propose a random S -matrix model for directed waves, based on the path integral solution to this Schrödinger equation. Exact computations are performed by summing over directed paths on the square lattice. We report asymptotic scaling laws for directed waves and interpret them in light of similar results for random walks.

Sums over paths are also used to investigate the properties of the 2D $\pm J$ Ising spin glass. We present an exact integer algorithm to compute the partition function of this system, based on the diagrammatic expansion of the high-temperature series. We examine the low-temperature behavior, the roots of the partition function in the complex plane, and the scaling laws for defects.

Boltzmann machines are neural networks based on Ising spin systems. We use exact techniques from statistical mechanics to derive an efficient learning algorithm for tree-like Boltzmann machines. The networks are shown to be capable of solving difficult problems in supervised learning.

Thesis advisor: Dr. Mehran Kardar
Professor of Physics

Acknowledgements

This thesis would not have been possible without the help and support of others. First and foremost, I would like to thank Mehran Kardar for supervising my graduate studies. I could not have asked for an advisor with greater clarity of thought or generosity of spirit. Countless are the times that I have benefited from his keen insights, devoted mentoring, and sage advice. I am also indebted to Nihat Berker and Lawrence Rosenson for serving on my thesis committee and for making several useful suggestions. Likewise, I wish to thank Michael Jordan for supporting my early efforts in the field of neural computation; I look forward to our future collaborations.

Peggy Berkovitz in the Graduate Physics Office and Imadiel Ariel in the Condensed Matter Theory Office have helped me on numerous occasions over the last four years. Thank you both for going above and beyond the call of duty.

I was fortunate to meet several kindred spirits during my time in graduate school. I am grateful for the friendship of fellow graduate students Hao Li, Deniz Ertas, and Daniel Lee, and my door will always be open to long-time office-mate and world-travelling companion, Leon Balents. To Tim Creamer, my most frequent and favorite Internet correspondent, I pledge to continue the email in the years to come.

I am deeply indebted to my parents and my brother Scott for their support and encouragement over the years. Finally, I wish to thank my wife Jacqueline for her love and support; you have made the last few years the happiest ones of all.

The work in this thesis was supported by an ONR graduate fellowship and by the NSF through grants DMR-93-03667 and PYI/DMR-89-58061.

Contents

1	Introduction	5
2	Directed Waves in Random Media	18
2.1	Introduction and Summary	18
2.2	Random S-Matrix Model	24
2.3	Analysis	38
3	The 2D $\pm J$ Ising Spin Glass	43
3.1	Introduction and Summary	43
3.2	An Exact Integer Algorithm	47
3.3	Results	55
3.3.1	Ground States and Low-Level Excitations	59
3.3.2	Roots of Z in the Complex Plane	65
3.3.3	Defects	71
3.4	Extensions	74
3.A	Periodic Boundary Conditions	78
3.B	Implementation	80
4	Learning in Boltzmann Trees	83
4.1	Introduction and Summary	83
4.2	Boltzmann Machines	88
4.3	Boltzmann Trees	93
4.4	Results	98
4.5	Extensions	100
5	Conclusion	103
	About The Author	114

Chapter 1

Introduction

Statistical mechanics connects the macroscopic properties of complex systems to the microscopic interactions of their constituent elements. Examples of such systems include the electrons in a copper wire, magnetic moments in a block of iron, and helium atoms in a sealed balloon. Due to the enormous number¹ of degrees of freedom in these systems, it is impossible to write down the true quantum-mechanical Hamiltonian and solve its Schrödinger equation, as one might do for the hydrogen atom. Instead, a starting point for understanding such systems is to formulate a simple, mathematical model that captures the important physics. Exact, perturbative, or approximate solutions of the model then make quantitative predictions that can be tested against experiment. This approach has historically met with a great deal of success. Thus, long before the advent of sophisticated tools such as the renormalization group, we had the Sommerfeld model[1] for electrons in normal metals, the mean-field theory of Weiss[2] for ferromagnets, and the Maxwell distribution[3] of molecular velocities for

¹typically, of order 10^{23} .

ideal gases. All these models led to significant advances, despite the fact that they did not take fully into account the interactions on the microscopic level.

The development of the renormalization group[4, 5] ushered in the modern era of statistical mechanics and attached a new importance to simple models of physical systems. The renormalization group uses scaling transformations to analyze the critical behavior of systems near a phase transition. It predicts that the physics at large length scales depends on only a few parameters, such as the dimensionality of the system and the type of symmetry breaking. One consequence of this is that many aspects of critical behavior do not depend on the detailed nature of microscopic interactions. This robustness explains the universal features of phase transitions that appear in very different sorts of systems—for example, uniaxial ferromagnets and liquid-gas mixtures. It also accounts for the remarkable success of phenomenological Landau theories[6] and idealized models.

A great deal of attention is now being focused on the statistical mechanics of disordered systems[7]. A characteristic feature of such systems is that they lack the translational symmetries of their pure counterparts. This makes the theoretical treatment of disordered systems more complicated. Of course, experimental realizations of disordered systems are ubiquitous, since it is rare that nature produces a perfect specimen of any material. Impurities, inhomogeneities, random perturbations—all of these can lead to interesting, new forms of behavior. In many systems, the disorder is essentially fixed or frozen on experimental time scales, unaffected by thermal noise even at high temperatures. The low-temperature (and sometimes $T = 0$) behavior of the rest of the system can be strongly influenced by the presence of such quenched

randomness. A recent development in the study of disordered systems has been the emergence of links to the field of neural computation[8]. Here, ideas from statistical mechanics have already helped to clarify a number of issues in optimization[9, 10, 11], associative memory[12], learning capacity[13] and generalization[14].

This thesis investigates exact computations in the statistical mechanics of disordered systems. The exactness and efficiency of our methods distinguish them from previous ones. Three different problems are discussed: directed waves in random media[15], the 2D $\pm J$ Ising spin glass[16], and Boltzmann learning in tree-like neural networks[17]. The first of these problems concerns the effect of a random potential on a zero-temperature dynamical system; the second and third deal with quenched randomness in Ising spin systems. This chapter introduces these problems and the ideas that unite them.

Many of the difficulties in analyzing disordered systems are best illustrated by example. The random walk is one of the oldest problems in probability theory[18], and one with several applications in physics[19]—Brownian motion, polymer chains, interfaces, and spin systems, to name only a few. Not surprisingly, the properties of random walks in pure and disordered systems are quite different. Consider, for example, a random walk on a simple cubic lattice in D dimensions. At time $t = 0$, the walker is situated at the origin of the lattice. At subsequent times $t = 1, 2$, etc., the walker randomly chooses to take a step in one of the $2D$ possible directions along the lattice.

In a pure system, the walker's decision does not depend on his location, and each direction is an equally likely choice for his next step. At large times t , the probability

$P(\mathbf{x}, t)$ to find the walker at site \mathbf{x} is given by the central limit theorem:

$$P(\mathbf{x}, t) = \frac{1}{(2\pi Dt)^{D/2}} \exp\left\{-\frac{\mathbf{x}^2}{2Dt}\right\}. \quad (1.1)$$

Many of the properties of random walks in pure systems follow immediately from this Gaussian probability distribution. Two important properties are the walker's mean-square displacement $\langle x^2 \rangle$ and the mean number of times t_0 that the walker returns to the origin. For long walks ($t \gg 1$), we have $\langle x^2 \rangle \sim t^{2\nu}$ with $\nu = 1/2$, and

$$t_0 \sim \begin{cases} t^{1/2} & \text{if } D = 1 \\ \ln t & \text{if } D = 2 \\ C_D & \text{if } D \geq 3, \end{cases} \quad (1.2)$$

with C_D a dimension-dependent constant. Note that in $D \leq 2$, the walker returns to the origin an infinite number of times as $t \rightarrow \infty$; this has led to the observation[18] that in $D \leq 2$, “all paths lead to Rome.” By contrast, the mean-square distance $\langle x^2 \rangle$ does not depend on dimension; it diffuses linearly in time for all values of D .

The situation is clearly different for random walks in disordered systems[20]. Anisotropies and inhomogeneities in the environment can modify the properties of random walks. For example, on a lattice with a large fraction of missing sites, the walker may remain localized around its starting point, even at large times. In this case, the walker will return to the origin many times more than predicted by eq. (1.2). Likewise, on a lattice with traps and delays, the walk may obey a subdiffusive scaling law, with $\langle x^2 \rangle \sim t^{2\nu}$ and $\nu < 1/2$. Other behaviors are also possible, depending on

the type of disorder.

The problem of random walks in disordered systems appears in various guises throughout this thesis. It first surfaces in Chapter II, where we investigate the unitary propagation of directed waves in random media. Such waves may be formed when a focused beam impinges on a medium with several scattering defects. In the limit of strong-forward scattering, these waves obey a Schrödinger equation for a quantum particle in a random time-dependent potential. We formulate a lattice model for directed waves based on the path integral solution to this Schrödinger equation. An attractive feature of our model is that the paths of directed wave fronts trace out random walks on an inhomogeneous lattice. The effect of inhomogeneities, and their relation to the disorder in the original problem, are discussed in detail. We look at two types of transverse fluctuations, $\overline{\langle x \rangle^2}$ and $\overline{\langle x^2 \rangle} - \overline{\langle x \rangle}^2$, where overbars denote averages over all possible landscapes of disorder. In the language of random walks, these quantities measure the typical deviation of the walker from the origin and the typical width of the walker's probability distribution. We show how to compute these quantities recursively as a function of time by summing over all possible random walks on the lattice. We also derive asymptotic scaling laws for them based on the properties of elementary random walks.

Another problem in statistical physics that can be mapped onto a sum over random walks is the two-dimensional (2D) Ising model[21, 22]. This model and its application to the study of disordered systems form the subject of Chapter III. The Ising model was originally introduced to study the phase transition in ferromagnetic solids such as iron and nickel. When samples of these materials are cooled below a critical

temperature T_c , they develop a non-zero magnetic moment, even in the absence of an applied magnetic field. While the value of T_c varies from material to material, many properties of the phase transition are universal to all ferromagnets. The understanding of these phase transitions, based on the theory of the renormalization group, is one of the triumphs of modern statistical mechanics.

The Ising model is an outstanding example of a simple, mathematical model that captures a great deal of important physics. The model illustrates how the spin-spin interactions between magnetic ions on a crystal lattice can lead to long-range, cooperative effects. In short-range Ising models, binary spins $S_i = \pm 1$ interact with their nearest neighbors on the lattice via pairwise couplings, or bonds, $\{J_{ij}\}$. In the presence of a magnetic field h , the Hamiltonian for the system is given by

$$\mathcal{H} = - \sum_{\langle ij \rangle} J_{ij} S_i S_j - \sum_i h S_i, \quad (1.3)$$

with the $\langle ij \rangle$ sum over all nearest-neighbor pairs on the lattice. The partition function $Z = \text{Tr}_S e^{-\mathcal{H}/T}$ encodes the thermodynamic behavior of the system at different values of the temperature T .

In a pure Ising ferromagnet, the bonds $J_{ij} = J > 0$ are uniform over the entire lattice. In this case, the ground state of the system, with all the spins aligned in parallel, exhibits long-range ferromagnetic order. In one-dimensional systems, this order is destroyed by thermal fluctuations, so that only at $T = 0$ does the system exhibit true long-range order. In two or more dimensions, however, the ferromagnetic order is stable to small thermal fluctuations, so that by gradually raising the temperature,

one encounters a finite-temperature transition between an ordered and disordered phase. Below $T < T_c$, in the ordered phase, the system exhibits a spontaneous magnetization, $\langle S_i \rangle \neq 0$, while above $T > T_c$, the spins fluctuate about a mean value of zero. The properties of phase transitions in ferromagnets are well-understood[5]. In the neighborhood of the critical point, the spin-spin correlations decay with distance as

$$\langle S_0 S_r \rangle - \langle S_0 \rangle \langle S_r \rangle \sim \frac{e^{-r/\xi}}{r^{d-2+\eta}},$$

where ξ is a temperature-dependent correlation length. As one approaches the critical point, the correlation length diverges algebraically as $\xi \sim |T - T_c|^{-\nu}$. Exactly at T_c , the spin-spin correlations have a power law decay, characterized by the critical exponent η . The values of ν and η are universal features of pure Ising ferromagnets; they depend on the dimension of the lattice, but are otherwise insensitive to microscopic details. Bulk quantities, such as the specific heat and magnetization, also exhibit singular behavior at phase transitions and have critical exponents associated with them. An important prediction of the renormalization group is that these other exponents are related to ν and η by simple scaling laws.

All the critical exponents are known exactly for the pure 2D Ising model, where the free energy and spin-spin correlations can be calculated analytically in closed form for zero magnetic field. A famous method for calculating the partition function of this system, due to Kac and Ward[23], is to recast the problem as a sum over random walks. The Kac-Ward method exploits the diagrammatic representation of terms in the high-temperature expansion as graphs on the square lattice. Kac and Ward

transformed the problem of evaluating these diagrams into a sum over random walks of different lengths. For the 2D Ising model with uniform ferromagnetic bonds, the sum is greatly simplified by the translational symmetries of the lattice. Kac and Ward performed the sum exactly for walks on an infinite 2D lattice and thus obtained the free energy for the pure Ising model in the thermodynamic limit. They also showed how to calculate free energies for systems of finite size.

An active area of research is to extend our understanding to phase transitions in disordered magnets, where entirely new types of behavior are found. Except for special cases, disordered magnets are not as well understood as their pure counterparts. Disordered magnets can be created in the laboratory by randomly diluting non-magnetic materials with magnetic ions. In certain situations, the exchange effects lead to the presence of competing ferromagnetic and antiferromagnetic interactions. Materials with these competing interactions are known spin glasses[24, 25], due to their slow relaxation times at low temperatures. Typical examples of spin glasses are CuMn and AuFe. At high temperatures, the local magnetic moments in spin glasses fluctuate about zero mean, but at low temperatures, they freeze in random directions. As in ferromagnets, these two regimes are believed to be separated by a phase transition.

The Ising model can also be used to study disordered magnets. To model the competing interactions in spin glasses with short-range interactions, the bonds J_{ij} that couple nearest-neighbor spins are chosen randomly from a probability distribution $p(J_{ij})$. Two common choices, introduced by Edwards and Anderson[26], are the

Gaussian distribution

$$p(J_{ij}) = \frac{1}{\sqrt{2\pi J^2}} e^{-J_{ij}^2/2J^2}$$

and the $\pm J$ distribution

$$p(J_{ij}) = \frac{1}{2}[\delta(J_{ij} - J) + \delta(J_{ij} + J)].$$

Edwards-Anderson models are believed to capture the important physics of Ising spin glasses, just as pure Ising models do for uniaxial ferromagnets. As in ferromagnets, one expects to find a diverging correlation length associated with the phase transition in spin glasses. In particular, just above T_c , it is believed that the mean-square correlations (averaged over pairs of spins) should behave as

$$\overline{\langle S_0 S_r \rangle^2} \sim \frac{e^{-r/\xi}}{r^{d-2+\eta}},$$

where the divergence of ξ and the power law decay are universal features of the model. Note that unlike the pure case, finding the ground state for the spin glass Hamiltonian is a non-trivial problem in optimization. Because it is impossible to satisfy the competing ferromagnetic and antiferromagnetic interactions, spin glasses are highly frustrated systems. We seek not only to understand the phase transition in these models, but also to characterize the properties of the low temperature phase.

In Chapter III, we apply the Kac-Ward method for computing partition functions to the 2D $\pm J$ spin glass. In particular, we present a polynomial-time algorithm to compute the exact integer density of states for spins on a $L \times L$ square lattice. The

algorithm uses the transition matrix proposed by Kac and Ward to sum over random walks of length $\ell \leq L^2$. Despite many previous studies which establish the occurrence of a zero-temperature phase transition in the 2D $\pm J$ spin glass, much remains to be understood about its low-temperature behavior. We investigate the divergence of the correlation length and the appearance of power-law correlations at the onset of the phase transition. We also examine the number of low-level excitations in the spin glass and the sensitivity to boundary conditions at zero temperature.

Though originally introduced to study phase transitions in magnets, Ising models have found a wide range of applications. Recently, they have been used to study models of parallel distributed processing[27, 28] and neural computation[8, 29]. The goal of these studies has been to understand the emergent computational abilities of networks of interconnected units. Inspired by the networks of neurons and synapses found in the brain[30], these connectionist networks have been trained to perform a number of tasks, including speech generation[31] and recognition[32], pattern classification[33], handwriting recognition[34], automobile navigation[35] and time series prediction[36]. Unlike traditional methods of computation, connectionist networks have the ability to learn from examples and typically exhibit a high degree of fault tolerance.

A simple example of such a network is the binary perceptron[37], shown in Figure 1.1. This network has one output unit S and N input units S_i connected by weights J_i and bias h . The units can take the values ± 1 and are related by

$$S = \text{sgn} \left(\sum_{i=1}^N J_i S_i + h \right).$$

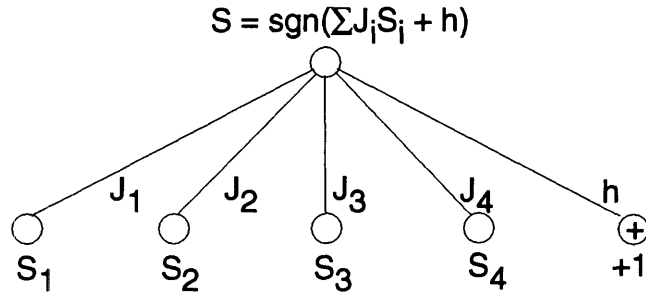


Figure 1.1: A binary perceptron with output S , inputs S_i , weights J_i , and threshold h . The output is equal to the sign of the weighted sum ($\sum J_i S_i + h$).

The binary perceptron thus computes the weighted sum of its input units and uses the sign for its output. The bias or offset h sets the threshold for the output unit to be positively activated; it can be interpreted as the effect of an additional input unit, fixed at the value $+1$. Clearly, the perceptron is severely restricted in the type of input-output mappings it can represent[38]. Given p_+ input patterns with output $+1$, and p_- input patterns with output -1 , a perceptron can perform the desired mapping only if the p_+ patterns are separated from the p_- patterns by a hyperplane in the N -dimensional input space. Not all problems are linearly separable in this way. Consider, for example, the problem of multiplication, or N -bit parity. The case of two input units ($N = 2$) is illustrated in Figure 1.2. Note that there is no line in the $S_1 S_2$ plane that separates the inputs whose product is $+1$ from those whose product is -1 . Consequently, no choice of weights will enable the perceptron to perform this mapping.

More expressive networks can be created by adding layers of hidden units between the input units and the output unit[28]. An example of a network with one layer of hidden units is shown in Figure 1.3. Ackley, Hinton, and Sejnowski[39] proposed

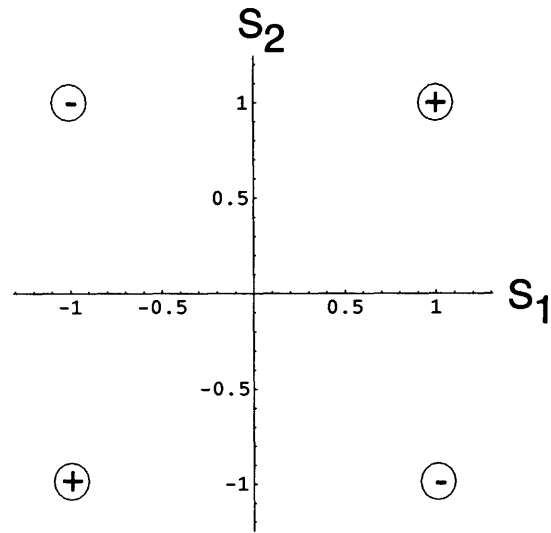


Figure 1.2: Linearly inseparability of the $N=2$ parity problem: it is impossible to draw a line that separates the $+1$ outputs from the -1 outputs.

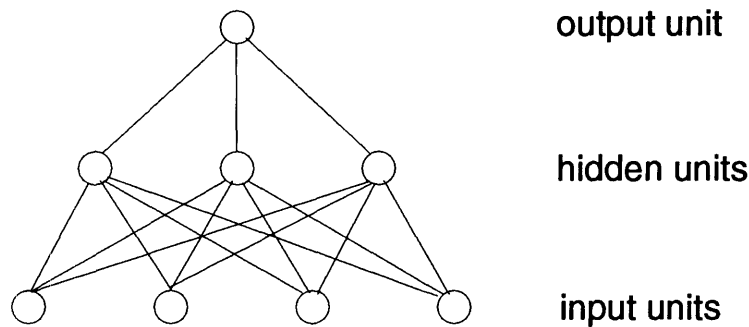


Figure 1.3: A Boltzmann machine with one layer of hidden units.

stochastic update rules for multi-layer networks based on the analogy to Ising spin systems. These rules guarantee that the equilibrium probability for the network to be found in state μ obeys the Boltzmann distribution $P_\mu = Z^{-1}e^{-\beta\mathcal{H}_\mu}$ where Z is the partition function for a system of Ising spins (or units) S_i , bonds (or weights) J_{ij} , and fields (or biases) h_i . The network is queried by clamping the input units to a particular pattern, then measuring the magnetization of the output unit. The input-output mapping performed by the network depends on the weights J_{ij} .

The Boltzmann learning algorithm prescribes how to adapt the weights in order to implement a desired mapping between input and output units. In this algorithm, individual weight changes ΔJ_{ij} are computed from the correlations $\langle S_i S_j \rangle$ between the i th and j th units in the network. In Chapter IV, we examine the properties of Boltzmann machines with tree-like connectivity between the output and hidden units. For networks of this type, we describe an efficient and economical way to implement the Boltzmann learning algorithm. Our method exploits the technique of decimation from statistical mechanics, originally developed in the context of the renormalization group. The strategy is similar to the one used to analyze complicated electric circuits, based on the rules for combining circuit elements in series and parallel. We present similar rules for combining weights in Boltzmann machines and use them to compute exact correlation functions in polynomial time.

The methods developed in this thesis for directed waves, the $2D \pm J$ spin glass, and Boltzmann machines can be applied to many other problems. Having demonstrated the potential for exact computations in these systems, we conclude in Chapter V by mentioning some areas for further research. The last chapter also issues some challenges for researchers with access to supercomputers and parallel processing machines. It is hoped that the implementation of our algorithms on these faster machines will lead to further insights into the statistical mechanics of disordered systems.

Chapter 2

Directed Waves in Random Media

2.1 Introduction and Summary

The problem of wave propagation in random media is one of longstanding theoretical interest[40, 41]. Only recently, however, have we begun to appreciate its connection to other problems in the physics of disordered systems, such as electron localization[42, 43], directed polymers in random media[44], and anomalous diffusion[20, 45]. Several authors[46, 47, 48] have suggested that the diffusion of *directed* wave fronts in disordered media is described, to a good approximation, by the Schrödinger equation for a particle in a random time-dependent potential. In this chapter, we propose a new model, based on random S -matrices, to explore the consequences of this description. An important aim of our study is to contrast the resulting behavior of waves with the types of diffusion known to occur in other disordered systems.

The approximations that reduce the full wave equation to the parabolic Schrö-

ding equation for directed waves[46] have been discussed most recently by Feng, Golubovic, and Zhang (FGZ)[48]. Here, we briefly review this reduction starting with the Helmholtz equation for propagation of a *scalar* wave Φ in a random medium. The static solution for Φ satisfies

$$[\nabla^2 + k^2 n^2(x, y, z)]\Phi(x, y, z) = 0, \quad (2.1)$$

where $n(x, y, z)$ is a nonuniform index of refraction that describes the landscape of disorder in the host medium. Following FGZ, we decompose $n^2(x, y, z) = n_o^2 + \delta n^2(x, y, z)$, where n_o is the disorder-averaged index of refraction, and $\delta n^2(x, y, z)$ contains local fluctuations due to randomly distributed scattering centers. The problem of directed waves arises in anisotropic media in which the scattering potential set up by these fluctuations varies slowly in the z direction, so as to favor coherent propagation along the z direction. For such a wave aimed parallel to the z axis, we can set $\Phi(x, y, z) = \Psi(x, y, z)e^{ikn_o z}$, thus reducing eq. (2.1) to

$$-\frac{\partial^2 \Psi}{\partial z^2} - 2ikn_o \frac{\partial \Psi}{\partial z} = \left(\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial y^2} \right) + k^2 \delta n^2(x, y, z)\Psi. \quad (2.2)$$

Wave propagation according to eq. (2.2) can be alternatively regarded as the scattering of photons by the fluctuations in n . We are interested in circumstances where the individual scattering events lead to a sequence of small fluctuations in the transverse momentum components of the z -directed paths. We would also like to ignore any back scattering, i.e. large changes in the longitudinal component of the photon

momentum. For these conditions to hold, we require $\delta n^2 \ll n^2$ and $\partial_z \delta n^2 \ll kn_o \delta n^2$. These conditions may be satisfied in anisotropic media[46, 47, 48] (e.g. with long fibers along the z -axis). The parabolic wave equation is thus obtained by ignoring the second derivative term on the left hand side of eq. (2.2). The analogy to the Schrödinger equation now becomes apparent, after the change of variable $z \leftrightarrow t$, which reduces eq. (2.2) to

$$i \frac{\partial \Psi}{\partial t} = [-\gamma \nabla^2 + V(x, y, t)] \Psi, \quad (2.3)$$

with $\gamma = (2kn_o)^{-1}$ and $V = -k\delta n^2/2n_o$. Eq. (2.3) appears in several contexts besides the problem of directed waves in random media. A quantum mechanical description of motion in *dynamically* disordered media has particular relevance for the problem of diffusion in crystals at finite temperature[45, 49, 50]. Random time-dependent potentials have also been used to model the environment of a light test particle in a gas of much heavier particles[51]. Thus although, as we shall discuss later, the applicability of eq. (2.3) to wave propagation in random media is somewhat limited, the study of its general scaling properties is of much intrinsic interest.

For generality, we examine the problem of directed waves in d dimensions. The solution to the appropriate Schrödinger equation is then given by the Feynman path-integral formula[46, 52, 53]

$$\Psi(\mathbf{x}, t) = \int_{(0,0)}^{(\mathbf{x},t)} \mathcal{D}\mathbf{x}(\tau) \exp \left\{ i \int_0^t d\tau \left[\frac{1}{2\gamma} \left(\frac{d\mathbf{x}}{d\tau} \right)^2 - V(\mathbf{x}(\tau), \tau) \right] \right\}, \quad (2.4)$$

where $\mathbf{x}(\tau)$ now describes a path in $d - 1$ dimensions. In writing eq. (2.4), we have chosen the standard initial condition that at time $t = 0$, the wave function is localized at the origin. The beam positions $\overline{\langle \mathbf{x}^2 \rangle}$ and $\overline{\langle \mathbf{x} \rangle^2}$ characterize the transverse fluctuations of the wave function Ψ about the forward path of least scattering. Here we use $\langle \dots \rangle$ to indicate an average with the weight $|\Psi(x, t)|^2$ for a given realization, and overbars to indicate quenched averaging over all realizations of randomness. Roughly speaking, $\overline{\langle \mathbf{x} \rangle^2}$ describes the wandering of the beam center, while $\overline{\langle \mathbf{x}^2 \rangle} - \overline{\langle \mathbf{x} \rangle^2}$ provides a measure of the beam width.

Path integrals similar to eq. (2.4) also appear in the two closely related problems of directed polymers (DP)[44] and strong localization[54, 55, 56]. In the former problem $\Psi(\mathbf{x}, t)$ represents the (positive) Boltzmann weight for the ensemble of DP configurations which connect the origin to (\mathbf{x}, t) : each path contributes an energy cost due to line tension, and a potential energy due to encounters with random impurities[44]. This problem is thus obtained by setting γ and $V(\mathbf{x}, \tau)$ imaginary in eq. (2.4). The quantum tunnelling probability of a strongly localized electron is also obtained by summing over all paths connecting the initial and final sites. In this case each path also acquires a random phase due to the effects of magnetic impurity scatterings[54]. This problem can thus be described by an imaginary γ , but a real V in eq. (2.4). We can thus pose the more general problem of studying the characteristic fluctuations of path integrals of the form eq. (2.4), when γ and V can take any values in the complex plane. Numerical and analytical evidence seems to indicate that DP and tunneling problems show similar scaling behaviors[54, 56]. We shall present some evidence indicating that the point corresponding to real γ and V in the complex plane,

i.e. representing directed waves, is the only point in this space that shows new scaling behavior for fluctuations.

A special property of eq. (2.3) which is valid only for real γ and V is *unitarity*, i.e. the norm $\int d\mathbf{x} |\Psi(\mathbf{x}, t)|^2$ is preserved at all times. (In the DP and tunnelling problems, the norm clearly decays exponentially with the length t .) This additional conservation law distinguishes the directed wave problem from DP and leads to a number of simplifications. Unitarity is of course a natural consequence of particle conservation for the Schrödinger equation, but it has no counterpart for directed wave propagation. It is likely that a beam of light propagating in a random medium will suffer a loss of intensity, due to either back-reflection, inelastic scattering, or localization phenomena[57].

Recent efforts to understand the diffusion of directed waves in random media have focused on the scaling behavior of the beam positions $\overline{\langle \mathbf{x}^2 \rangle}$ and $\overline{\langle \mathbf{x} \rangle^2}$ at large t . *Lattice* models have been used here with some success. It has been shown using density-matrix techniques, for instance, that $\overline{\langle \mathbf{x}^2 \rangle}$ scales linearly in time as a consequence of unitarity[49]; recent numerical simulations[58, 59] also support this view. The scaling behavior of $\overline{\langle \mathbf{x} \rangle^2}$ at large t , however, has proved more controversial. The first numerical work in this area was done by FGZ[48], who used a discretization procedure in which the norm of the wave function was *not* strictly preserved. In $2d$, they found that $\overline{\langle \mathbf{x} \rangle}$ grew superdiffusively as t^ν with $\nu \approx \frac{3}{4}$, while in $3d$, they found a phase transition separating regimes of weak and strong disorder. Recent numerical studies[58, 59, 60] on directed waves in $2d$ cast doubt on the validity of these results when the time evolution is strictly unitary. These studies report that $\overline{\langle \mathbf{x} \rangle^2}$ scales

subdiffusively in $2d$ as $t^{2\nu}$ with $\nu \approx 0.25 - 0.30$. Bouchaud et al[59] also conjecture that the wave function becomes “multifractal” in that an infinite number of critical exponents are required to describe its evolution.

Somewhat surprising is the fact that a *continuum* formulation of the wave problem leads to different results. An exact treatment of the continuum Schrödinger equation (2.3) has been given by Jayannavar and Kumar[50]. They show that for a random potential δ -correlated in time, $\overline{\langle \mathbf{x}^2 \rangle} \sim t^3$ as $t \rightarrow \infty$. This behavior is modified when there are short-range correlations in time[51], but the motion remains non-diffusive in that the particle is accelerated indefinitely as $t \rightarrow \infty$. Lattice models introduce a momentum cutoff $p_{max} \sim a^{-1}$, where a is the lattice spacing, and therefore do not exhibit this effect. The momentum cutoff generated by the lattice discretization is in some sense artificial. Nevertheless, in a real fluctuating medium, we do expect on large time scales to recover the lattice result, i.e. normal diffusion. The reason is that dissipative effects do generate an effective momentum cutoff in most physical systems. (Strictly speaking, even in the absence of dissipation, relativistic constraints lead to a velocity cutoff $v = c$.) The presence of such a cutoff for the wave propagation problem, and hence the physical relevance of lattice versus continuum models, is still a matter of debate. While there is no underlying lattice, one suspects on physical grounds that there does exist an effective momentum cutoff for propagating waves, related to the speed of light in the background medium.

In this study, we investigate a new model for the propagation of directed waves in strongly disordered multiple-scattering media. Our model is formulated on a discrete lattice and reproduces the result that the beam position $\overline{\langle \mathbf{x}^2 \rangle}$ grows linearly in time.

We find also that $\overline{(\mathbf{x})^2}$ scales as $t^{2\nu}$ with $\nu = \frac{1}{4}$ in $2d$ and as $\ln t$ in $3d$. Our approach is novel in several respects. First, our model is formulated in such a way that unitarity is manifestly preserved in numerical simulations, without resort to complicated checks. Second, we implement scattering events in a manner consistent with the local conservation of probability flux. Third, we perform all averages over disorder *exactly*, whereas previous studies resort to averaging over a necessarily finite number of computer-generated random environments. Finally, we look at scaling behavior in systems that are an order of magnitude larger than those previously considered.

The rest of the paper is divided into two parts. In Section II, we develop our model in considerable detail, with emphasis on the simplifying features that permit one to compute averages over disorder exactly. At the end of this section, we present the results of our $2d$ and $3d$ numerical simulations. Then, in Section III, we interpret our results in light of well-known properties of random walks. We conclude with some final comments on the connection to the DP problem.

2.2 Random S-Matrix Model

Previous numerical investigations of the problem have started by rewriting the Schrödinger equation (2.3) as a difference equation. Such an approach has the advantage of reducing straightforwardly to the continuum description as the unit time increment is shrunk to zero. Unfortunately, the naive discretization of eq. (2.3) does not preserve the unitarity of time evolution. Since most evidence suggests that it is precisely the constraint of unitarity that gives rise to a new universality class for directed waves,

this breakdown is quite serious. Realizing this, previous workers have enhanced the above discretization in ways to mitigate the breakdown of unitarity[48, 58, 59, 60]. We take a different approach and look for a discretization that manifestly preserves unitarity.

The fundamental motivation for our approach is the path integral description of quantum mechanics. Rather than discretizing the wave equation (2.3), we seek to implement the sum-over-histories prescription of the path integral (2.4). To this end, let us consider the general problem of a quantum particle on a spacetime lattice, initially localized at point A. We propose to assign a complex-valued amplitude to each particle trajectory on the lattice that emanates from A. Additionally, we want to impose the physical requirement that the probability current of the particle satisfies a local conservation law. The normalized wavefunction of the particle at point B can then be computed by summing the amplitudes of all trajectories that connect A to B. The number of these trajectories is finite due to the discretization of spacetime. We now show that the sum-over-histories approach, combined with the requirement of probability conservation, gives rise to a model in which the unitarity of time evolution is manifestly preserved.

For concreteness we introduce the model in $2d$. A discussion of its generalization to higher dimensions is taken up later. As is customary in the study of directed waves, we identify the time axis with the primary direction of propagation. Our first step, then, is to consider diffusion processes on the $2d$ lattice shown in Figure 2.1. It is amusing to note that this lattice has also been used for the discretizing the path integral of a relativistic particle in one dimension[52].

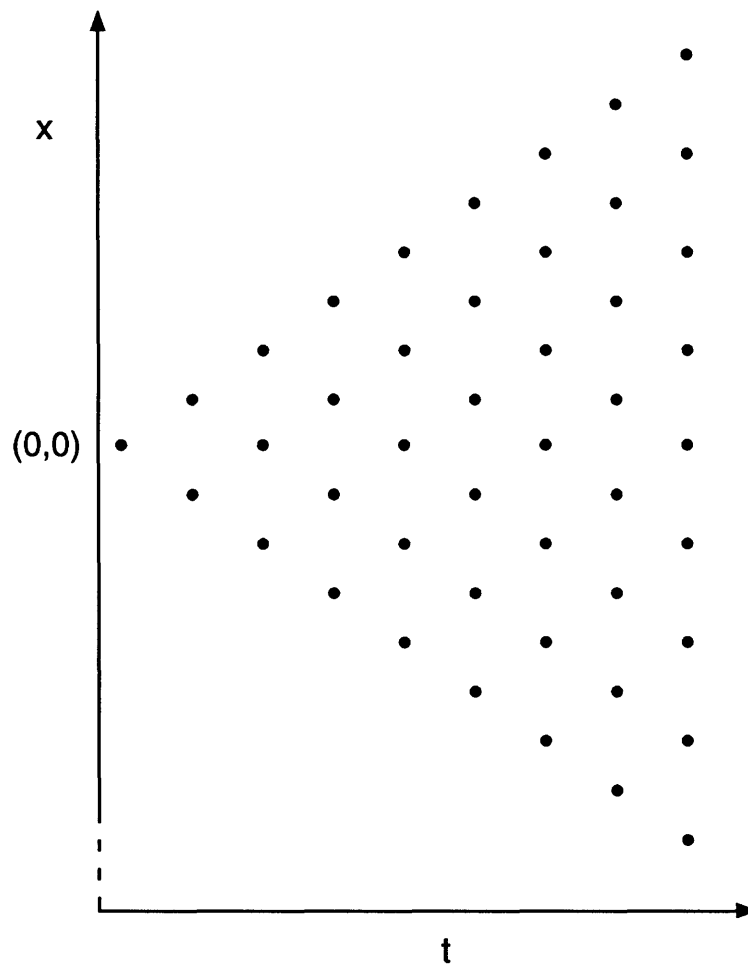


Figure 2.1: Lattice discretization for directed waves in $d = 2$. The wave function $\Psi_{\pm}(x, t)$ is defined on the links of the lattice, while random scattering events occur at the sites.

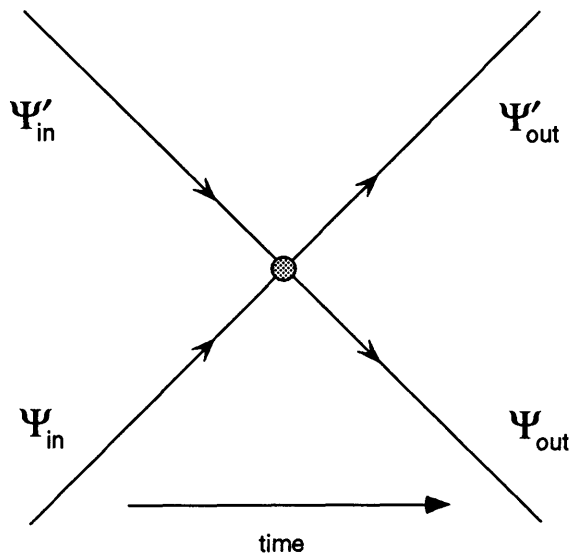


Figure 2.2: Scattering event at a lattice site. Time flows in the horizontal direction. A 2×2 S -matrix relates ingoing and outgoing amplitudes.

The wave function in our approach takes its values on the links of this lattice. We use $\Psi_{\pm}(x, t)$ to refer to the amplitude for arriving at the site (x, t) from the $\pm x$ direction (see Figure 2.2). At $t = 0$, the wave function is localized at the origin, with $\Psi_{\pm}(0, 0) = 1/\sqrt{2}$. Following the sum-over-histories prescription, our next step is to assign a complex-valued amplitude to each trajectory on the lattice emanating from the origin. Transfer matrix techniques lend themselves naturally to this purpose. To each site on the lattice, we therefore assign a 2×2 unitary matrix $S(x, t)$. The values of the wave function at time $t + 1$ are then computed from the recursion relation:

$$\begin{bmatrix} \Psi_+(x-1, t+1) \\ \Psi_-(x+1, t+1) \end{bmatrix} = \begin{bmatrix} S_{11}(x, t) & S_{12}(x, t) \\ S_{21}(x, t) & S_{22}(x, t) \end{bmatrix} \begin{bmatrix} \Psi_+(x, t) \\ \Psi_-(x, t) \end{bmatrix}. \quad (2.5)$$

The S -matrices are required to be unitary in order to locally preserve the norm of the wave function.

The S -matrix procedure outlined above weights each trajectory on the lattice with a complex amplitude. Consider, for example, the trajectory in which the particle, incident at the origin from the $-x$ direction, takes two steps in the $+x$ direction then two steps back. The amplitude \mathcal{A} assigned to this trajectory is given by the product of S -matrix elements:

$$\mathcal{A} = S_{21}(0, 0)S_{21}(1, 1)S_{22}(2, 2)S_{22}(1, 3). \quad (2.6)$$

In general, a trajectory of L links on the lattice is weighted with an amplitude derived from the product of L S -matrix elements. The value of the wavefunction $\Psi_{\pm}(x, t)$ is obtained by summing the individual amplitudes of all directed paths which start at the origin and arrive at the point (x, t) from the $\pm x$ direction. To simulate the effect of a random potential, we choose the S -matrices randomly from the group of 2×2 unitary matrices. We thus achieve a unitary discretization of the path integral in eq. (2.4), in which the phase change from the random potential $V(x, t)$ is replaced by an element of the matrix $S(x, t)$. The recursion relation in eq. (2.5) is the coarse-grained analogue of the Schrödinger equation (2.3); unlike a simple difference equation, however, eq. (2.5) enforces the local conservation of probability flux and leads to a sum-over-histories solution for the wavefunction. Unitarity is manifestly preserved.

Besides these advantages, the S -matrix approach also has a natural physical interpretation for the problem of directed waves in random media. The basic idea is simple: at time t , we imagine that a random scattering event occurs at each site in

the lattice at which either $\Psi_+(x, t)$ or $\Psi_-(x, t)$ is non-zero. The matrices $S(x, t)$, which relate the ingoing and outgoing amplitudes at each lattice site, can then be regarded as scattering matrices in the usual sense. Figure 2.2 illustrates a typical scattering event. A lattice S -matrix approach for the study of electron localization and the quantum Hall effect has been used by Chalker and coworkers[61]. A related model has also been recently proposed[62] to investigate the localization of wave packets in random media. These models also include back scattering and hence involve a larger matrix at each site.

We are interested in the beam positions

$$\overline{\langle x^2 \rangle}_t = \sum_x \overline{P(x, t)} x^2, \quad (2.7)$$

and

$$\overline{\langle x \rangle}_t^2 = \sum_{x_1, x_2} \overline{P(x_1, t) P(x_2, t)} x_1 x_2. \quad (2.8)$$

Here, $P(x, t)$ is the probability distribution function (PDF) on the lattice at time t , defined by:

$$P(x, t) = |\Psi_+(x, t)|^2 + |\Psi_-(x, t)|^2. \quad (2.9)$$

(Defining the weights directly on the bonds does not substantially change the results.)

Note that unlike the DP problem, $P(x, t)$ is properly normalized, i.e.

$$\sum_x P(x, t) = 1$$

and eqs. (2.7) and (2.8) are not divided by normalizations such as $\sum_x P(x, t)$. This simplification, a consequence of unitarity, is what makes the directed wave problem tractable.

The disorder-averages in eqs. (2.7) and (2.8) are to be performed over a distribution of S -matrices that closely resembles the corresponding distribution for V in the continuum problem. However, by analogy to the DP problem[44], we expect any disorder to be relevant. Hence, to obtain the asymptotic scaling behavior, we consider the extreme limit of strong scattering in which each matrix $S(x, t)$ is an *independently* chosen, random element of the group $U(2)$. With such a distribution we lose any pre-asymptotic behavior associated with weak scattering[51]. The results are expected to be valid over a range of length scales $a \ll x \ll \xi$, where a is a length over which the change of phase due to randomness is around 2π , and ξ is the length scale for the decay of intensity and breakdown of unitarity. Since the parabolic wave equation was obtained from the full wave equation (2.1) by assuming that the scattering potential varied slowly along the propagation direction ($\partial_z \delta n^2 \ll kn_o \delta n^2$), it is fair to inquire if the conditions for the validity of such path integrals are ever satisfied in transmission of light. As a partial answer, we provide an idealized *macroscopic* realization in which a beam of light is incident upon a lattice of beam splitters arranged as in Figure 2.2. Each splitter partially reflects and partially transmits the beam, both in the forward direction. (Note that as long as the beam width is smaller than the size of each slab, the beam does not encounter variations of n along the t direction, and will not be backscattered.) In this strong scattering limit, the effect of an impurity at (x, t) is therefore to redistribute the incident probability flux $P(x, t)$ at random in

the $+x$ and $-x$ directions. On average, the flux is scattered symmetrically so that the disorder-averaged PDF describes the event space of a classical random walk:

$$\overline{P(x,t)} = \frac{t!}{\left(\frac{t-x}{2}\right)! \left(\frac{t+x}{2}\right)!}. \quad (2.10)$$

Substituting this into eq. (2.7), we find $\overline{\langle x^2 \rangle}_t = t$, in agreement with previous studies[49]. Consider now the position of the beam center $\overline{\langle x \rangle}_t^2$, given by eq. (2.8). Unlike $\overline{P(x,t)}$, the correlation function $\overline{P(x_1,t)P(x_2,t)}$ does not have a simple form. An exact calculation of $\overline{\langle x \rangle}_t^2$ thus proves rather difficult.

One way to proceed is to perform numerical simulations, based on eq. (2.5), in which averages over disorder are computed by sampling a finite number of computer-generated random environments. For the purpose of computing $\overline{\langle x \rangle}_t^2$, however, this S -matrix algorithm has a large amount of unnecessary overhead. All the information required to compute beam positions is contained in the function $P(x,t)$. Moreover, we are not interested in those quantities, such as transverse probability currents, for which a complete knowledge of $\Psi_{\pm}(x,t)$ is required. A better algorithm, for our purposes, would be one that directly evolves $P(x,t)$ rather than the wave functions $\Psi_{\pm}(x,t)$.

One may wonder if such an algorithm exists, since in general, it is not possible to simulate the dynamics of the Schrödinger equation without reference to the wave function. Consider, however, the scattering event shown in Fig. 2.2. Probability flux

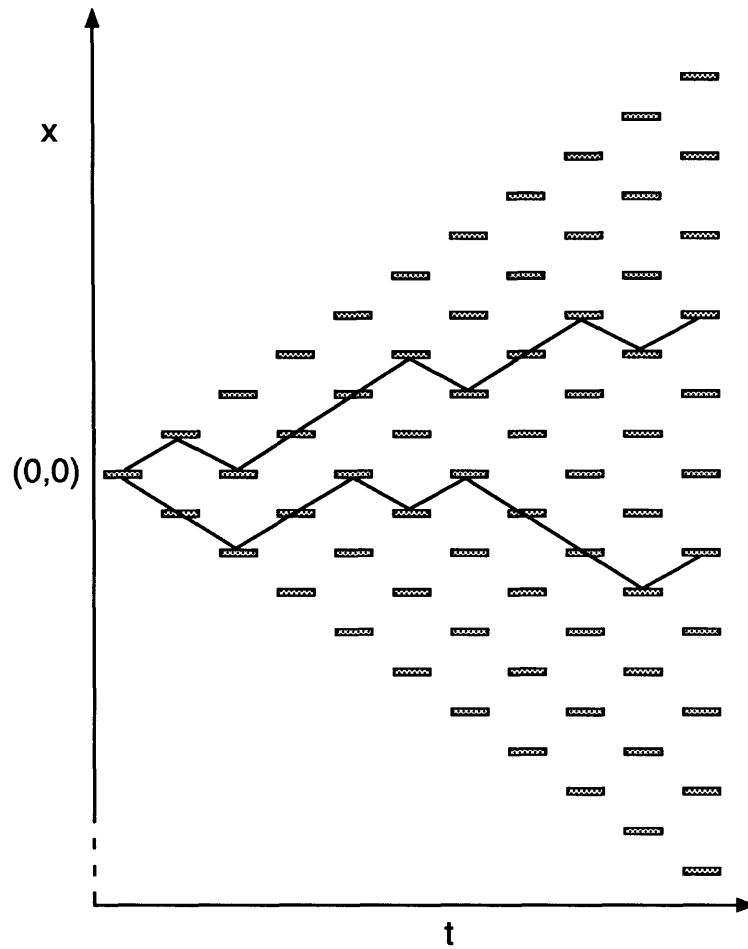


Figure 2.3: Lattice of beam splitters in $d = 2$. In black: a pair of paths contributing to $W(r, t)$, the disorder-averaged probability that two paths are separated by $2r$ at time t .

is locally conserved; hence,

$$|\Psi_i|^2 + |\Psi'_i|^2 = |\Psi_o|^2 + |\Psi'_o|^2. \quad (2.11)$$

As the S -matrix that connects these waves is uniformly distributed over the group $U(2)$, its action distributes the outgoing waves uniformly over the set of spinors whose components satisfy eq. (2.11). A straightforward calculation shows in turn that the ratio

$$\kappa = \frac{|\Psi_o|^2}{|\Psi_i|^2 + |\Psi'_i|^2} \quad (2.12)$$

is uniformly distributed over the interval $[0, 1]$. This result, which holds for all scattering events on the lattice, can be used to evolve $P(x, t)$ directly, without reference to the wave functions $\Psi_{\pm}(x, t)$.

Let us examine in detail how this is done. At $t = 0$, $P(x, t)$ is localized at the origin:

$$P(x, t = 0) = \delta_{x,0}. \quad (2.13)$$

As before, we imagine that at times $t > 0$, disorder-induced scattering events occur at all sites on the lattice where $P(x, t)$ is non-zero. Now, however, we implement these events by assigning to each lattice site a random number $0 \leq \kappa(x, t) \leq 1$. The probability distribution function $P(x, t)$ is then directly evolved according to the recursion relation

$$P(x, t + 1) = \kappa(x - 1, t)P(x - 1, t) + \{1 - \kappa(x + 1, t)\}P(x + 1, t). \quad (2.14)$$

When the numbers κ in eq. (2.14) are distributed uniformly between 0 and 1, this set of rules for evolving $P(x, t)$ is equivalent to the previous one for evolving $\Psi_{\pm}(x, t)$. In other words, calculating $\overline{\langle x \rangle_i^2}$ by updating $P(x, t)$ and averaging over $\kappa(x, t)$ is the same as calculating $\overline{\langle x \rangle_i^2}$ by updating $\Psi_{\pm}(x, t)$ and averaging over $S(x, t)$. In fact, we will see later that except in very special circumstances, eq. (2.14) leads to the same scaling behavior as long as $\bar{\kappa} = 1/2$.

So far, then, we have sketched two algorithms that can be used to investigate the scaling behavior of $\overline{\langle x \rangle_i^2}$. The first method evolves the wave functions $\Psi_{\pm}(x, t)$ through a field of random S -matrices. The second evolves the PDF $P(x, t)$ directly, with much less overhead. The *exact* equivalence of these two methods depends crucially on our choice of a uniform distribution for the S -matrices that appear in eq. (2.5). If the S -matrices are not chosen from a uniform distribution over the group $U(2)$, then the ratio κ defined by eq. (2.12) will not be distributed over the interval $[0, 1]$ in the same way at all lattice sites. Moreover, a non-universal distribution for κ invalidates the logic behind eq. (2.14). We emphasize, however, that the scaling behavior of $\overline{\langle \mathbf{x} \rangle^2}$ should not depend sensitively on the details of the distribution used to generate the S -matrices in eq. (2.5); a broad range of distributions should belong to the same universality class of diffusive motion. Consequently, the simplifying assumption of a uniform distribution should not destroy the generality of the results for directed waves in random media and/or quantum mechanics in a random time-dependent potential. The second method thus retains the essential elements of the problem, while from a computational point of view, it is greatly to be preferred.

In fact, the greatest virtue of the latter method is that it permits an even further

simplification. Indeed, though faster, more efficient, and conceptually simpler, it still suffers an obvious shortcoming: the averages over disorder are performed by sampling only a finite number of computer-generated realizations of randomness. We now show how to compute these averages in an exact way.

Define the new correlation function

$$W(r, t) = \sum_x \overline{P(x, t)P(x + 2r, t)}. \quad (2.15)$$

From eq. (2.13), we have the initial condition

$$W(r, t = 0) = \delta_{r,0}. \quad (2.16)$$

The value of $W(r, t)$ is the disorder-averaged probability that two paths, evolved *in the same realization of randomness*, are separated by a distance $2r$ at time t . We can compute this probability as a sum over all pairs of paths that meet this criteria. A typical configuration of paired paths is shown in Fig. (2.2). Consider now the evolution of two such paths from time t to time $t + 1$. Clearly, at times when $r \neq 0$, the two paths behave as independent random walks. On the other hand, when $r = 0$, there is an increased probability that the paths move together as a result of participating in the same scattering event. These observations lead to a recursion

relation for the evolution of $W(r, t)$:

$$W(r, t+1) = \left[\frac{1 + \Delta \delta_{r,0}}{2} \right] W(r, t) + \left[\frac{1 - \Delta \delta_{r,1}}{4} \right] W(r-1, t) + \left[\frac{1 - \Delta \delta_{r,-1}}{4} \right] W(r+1, t), \quad (2.17)$$

with $\Delta = 4(\overline{\kappa^2} - \bar{\kappa}^2) \geq 0$. The value of Δ measures the tendency of the paths to stick together on contact. As mentioned before, a uniform distribution of S -matrices over $U(2)$ gives rise to a uniform distribution of κ over the interval $[0, 1]$. In this case, $\Delta = 1/3$.

Starting from eq. (2.17), we have found $W(r, t)$ numerically for various values of $0 < \Delta < 1$. The position of the beam center was then calculated from

$$\overline{\langle x \rangle_t^2} = t - 2 \sum_r W(r, t) r^2. \quad (2.18)$$

The results for $t < 15000$, shown in Figure 2.2, suggest unambiguously that $\overline{\langle x \rangle_t^2}$ scales as $t^{2\nu}$, with $\nu = 1/4$. We emphasize here the utility of the S -matrix model for directed waves in random media. Not only does our final algorithm compute averages over disorder in an exact way, but it requires substantially less time to do so than simulations which perform averages by statistical sampling. We have in fact confirmed our $2d$ results with these slower methods on smaller lattices ($t < 2000$).

We now consider the S -matrix model in higher dimensions. Most of the features of the $2d$ model have simple analogues. The wave function takes its values on the links of a lattice in d dimensions. Random $N \times N$ S -matrices are then used to simulate scattering events at the sites of the lattice. The value of N is equal to one-half

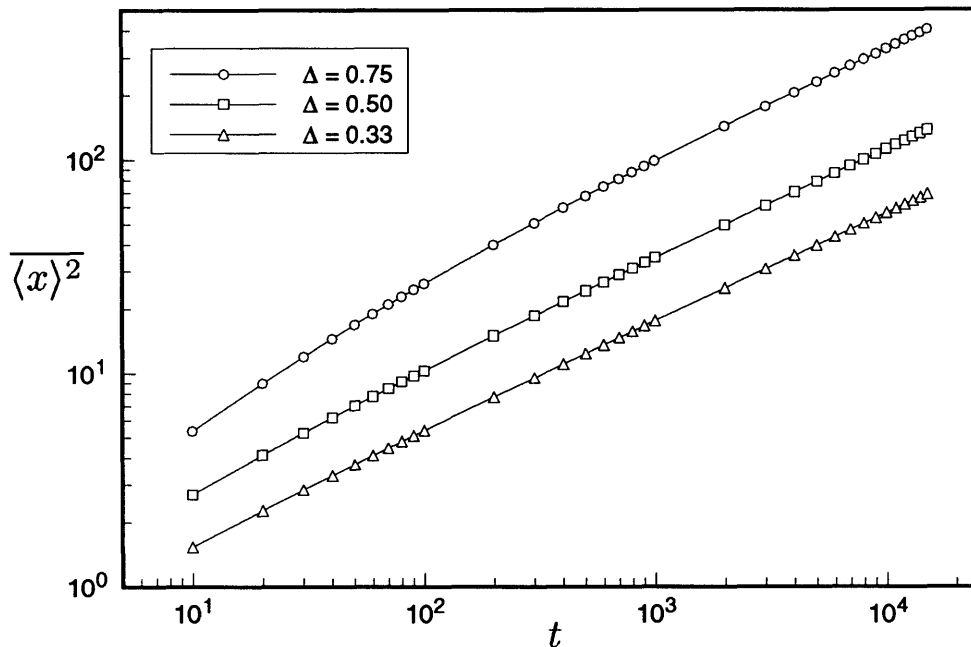


Figure 2.4: Log-log plot of the wandering of the beam center $\overline{\langle \mathbf{x} \rangle^2}$ versus the propagation distance t in $d = 2$, for various values of Δ (see eq. (2.17)).

the coordination number of the lattice. When the matrices $S(\mathbf{x}, t)$ are distributed uniformly over the group $U(N)$, the same considerations as before permit one to perform averages over disorder in an exact way. In addition, one obtains the general result for $d \geq 2$ that $\overline{\langle \mathbf{x}^2 \rangle}$ scales linearly in time.

The computation of $\overline{\langle \mathbf{x} \rangle^2}$ in $d > 2$, of course, requires significantly more computer resources. In $3d$, methods which rely on sampling a large number of realizations of randomness begin to lose their practical value. We have computed $\overline{\langle \mathbf{x} \rangle^2}$ on a $3d$ body-centered cubic lattice, starting from the appropriate generalization of eq. (2.17). The results for $t < 3000$, shown in Figure 2.2, indicate that $\overline{\langle \mathbf{x} \rangle^2}$ scales logarithmically in time.

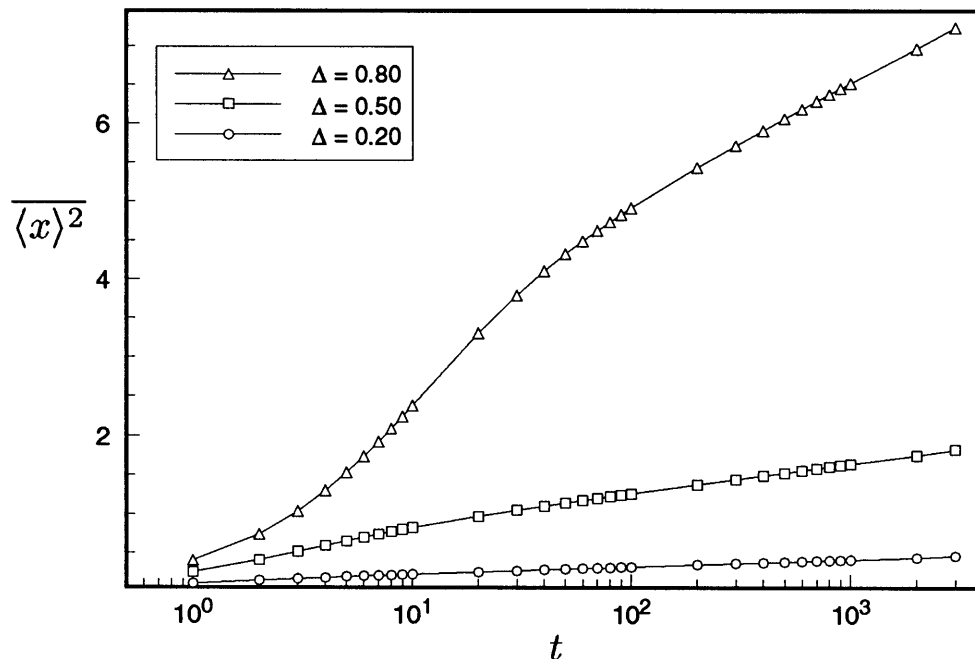


Figure 2.5: Semi-log plot of the wandering of the beam center $\overline{\langle \mathbf{x} \rangle^2}$ versus the propagation distance t in $d = 3$, for various values of Δ (see eq. (2.17)).

2.3 Analysis

In this section, we examine our numerical results in light of well-known properties of random walks. Consider a random walker on a simple cubic lattice in $D = d - 1$ dimensions. We suppose, as usual, that the walker starts out at the origin, and that at times $t = 0, 1, 2, \dots$, the walker has probability $0 < p \leq \frac{1}{2D}$ to move one step in any lattice direction and probability $1 - 2Dp$ to pause for a rest. As pointed out in the introduction to this thesis, the mean time t_0 spent by the walker at the origin grows as:

$$t_0 \sim \begin{cases} t^{1/2} & \text{if } D = 1 \\ \ln t & \text{if } D = 2 \\ C_D & \text{if } D \geq 3, \end{cases} \quad (2.19)$$

with C_D a dimension-dependent constant. From the numerical results of the previous section, it is clear that the same scaling laws describe the wandering of the beam center, $\overline{\langle \mathbf{x} \rangle^2}$, in $d = D + 1$ dimensions, for $d = 2$ and $d = 3$. We now show that this equivalence is not coincidental; moreover, it strongly suggests that $d_u = 3$ is a critical upper dimension for directed waves in random media.

To this end, let us return to our model for directed waves in $d = 2$. Applying the recursion relation for $W(r, t)$, eq. (2.17), to the identity for the beam center, eq. (2.18), gives

$$\begin{aligned} \overline{\langle x \rangle_t^2} - \overline{\langle x \rangle_{t-1}^2} &= 1 - 2 \sum_r [W(r, t) - W(r, t-1)] r^2 \\ &= \Delta W(0, t-1). \end{aligned} \tag{2.20}$$

Summing both sides of this equation over t , one finds

$$\overline{\langle x \rangle_t^2} = \Delta \sum_{\tau=0}^{t-1} W(0, \tau). \tag{2.21}$$

In the previous section, we saw that the disorder-averaged correlation function $W(r, t)$ describes the time evolution of two paths in the same realization of randomness. We can also regard $W(r, t)$ as a probability distribution function for the relative coordinate between two interacting random walkers. In this interpretation, the value of Δ in eq. (2.17) parametrizes the strength of a contact interaction between the walkers. If $\Delta = 0$, the walkers do not interact at all; if $\Delta = 1$, the walkers bind on contact.

According to eq. (2.21), the wandering of the beam center $\overline{\langle x \rangle_t^2}$ is proportional to the mean number of times that the paths of these walkers intersect during time t . If $\Delta = 0$, the number of intersections during time t obeys the scaling law in eq. (2.19), since in this case, the relative coordinate between the walkers performs a simple random walk. Our numerical results indicate that the same scaling law applies when $0 < \Delta < 1$: the contact attraction does *not* affect the asymptotic properties of the random walk. To elaborate this point, we expand $W(r, t)$ as a power series in Δ :

$$W(r, t) = \sum_{n=0}^{\infty} \Delta^n W_n(r, t).$$

The zeroth order term in this series, $W_0(r, t)$, describes a simple random walk, while higher order terms represent corrections due to the contact attraction Δ . Substituting into eq. (2.21) and using the $D = 1$ result of eq. (2.19) give

$$\overline{\langle x \rangle_t^2} \sim \Delta t^{1/2} \left[1 + \mathcal{O}\left(\frac{\Delta}{t^{1/2}}\right) \right]. \quad (2.22)$$

The scaling properties of higher-order corrections follow from simple dimensional arguments: from eq. (2.17), it is apparent that Δ has units of $[x]$, since it multiplies a Kronecker delta function. Noting that in the continuum limit, eq. (2.17) becomes a diffusion equation, we also have that $[x] = [t]^{1/2}$, so that higher-order corrections must be smaller by a relative factor of $\Delta t^{-1/2}$. The series thus converges rapidly for large t , and we conclude that $\nu = 1/4$ exactly in $d = 2$.

The above argument is readily generalized to $d > 2$, in which Δ has the units

of $[x]^{d-1} = [t]^{(d-1)/2}$. The result is that the wandering of the beam center, $\overline{\langle \mathbf{x} \rangle^2}$, in $d = D + 1$ dimensions obeys the scaling laws in eq. (2.19), with next order corrections smaller by relative factors of $\mathcal{O}(\Delta/t^{\frac{d-1}{2}})$. Moreover, the argument leads to an upper critical dimension $d_u = 3$ above which the typical wandering of the beam center remains finite even as the propagation distance $t \rightarrow \infty$. In summary, three classes of behavior are thus encountered in this model. For $\Delta = 0$, i.e. no randomness, the incoming beam stays centered at the origin, while its width grows diffusively. For $0 < \Delta < 1$, the beam center, $\overline{\langle \mathbf{x} \rangle^2}$, also fluctuates, but with a dimension dependent behavior as in eq. (2.19). In the limit of $\Delta = 1$, interference phenomena disappear completely. (This limit can be obtained by replacing the beam splitters of Figure 2.2 with randomly placed mirrors.) In this case, the beam width is zero, and the beam center performs a simple random walk.

To conclude, we compare the situation here to the one of directed polymers in random media[44]. In the replica approach to the DP problem, the n -th moment of the weight $\Psi(x, t)$ is obtained from the statistics of n directed paths. Disorder-averaging again produces an attractive interaction between these paths, with the result that the paths can be regarded as the world lines of n quantum particles interacting through a short-range potential. The large t behavior of n -th order moments is then related to the ground state wave function of the corresponding n -body problem in $d - 1$ dimensions. In $d = 2$, the Bethe ansatz can be used to find an exact solution for particles interacting through delta function potentials: any amount of randomness (and hence attraction) leads to the formation of a bound state. The behavior of the bound state energy can then be used to extract an exponent of $\nu = 2/3$ for the

superdiffusive wandering of the single DP in the quenched random potential.

By contrast, the replicated paths encountered in the directed wave problem (such as the two paths considered for eq. (2.15), although interacting, cannot form a bound state. This point was first emphasized by Medina et al[58], who showed that the formation of a bound state was inconsistent with the constraints imposed by unitarity on the lattice. This result also emerges in a natural way from our model of directed waves. In $d = 2$, for instance, it is easy to check that $W(\mathbf{r}, t) \sim (1 - \Delta\delta_{r,0})^{-1}$ is the eigenstate of largest eigenvalue for the evolution of the relative coordinate. Hence, as $t \rightarrow \infty$, for randomness δ -correlated in space and time, there is no bound state. This result holds in $d \geq 2$ and is not modified by short-range correlations in the randomness. The probability-conserving nature of eq. (2.17) is crucial in this regard. Small perturbations that violate the conservation of probability lead to the formation of a bound state. In the language of the renormalization group, this suggests that the scaling behavior of directed waves in random media is governed by a fixed point that is unstable with respect to changes that do not preserve a strictly unitary time evolution. Numerical and analytic results support the idea that this fixed point belongs to a new universality class of diffusive behavior.

Acknowledgements

The work in this chapter represents a collaboration with Mehran Kardar and Nick Read. Useful discussions with E. Medina and H. Spohn are gratefully acknowledged.

Chapter 3

The 2D $\pm J$ Ising Spin Glass

3.1 Introduction and Summary

The last fifteen years have witnessed a great deal of work on spin glasses[9, 24, 25, 63]. Nevertheless, the description of the phase transition and the nature of the ordered state remain controversial subjects[64, 65]. The starting point for most theoretical work is the Edwards-Anderson (EA) Hamiltonian[26]

$$\mathcal{H} = - \sum_{ij} J_{ij} \sigma_i \sigma_j, \quad (3.1)$$

where the J_{ij} are quenched random variables and the σ_i are Ising spins on a regular lattice. Interactions with infinite-range[66] lead to a saddle point solution with broken replica symmetry[9]. It is not known, however, to what extent this mean-field result captures the behavior of short-range interactions[65, 67].

A widely studied model is the $\pm J$ spin glass[68], in which the sign of each bond is

random but its magnitude fixed. In two dimensions, the $\pm J$ spin glass with nearest-neighbor interactions exhibits a phase transition at zero temperature[69]. The properties of this $T = 0$ transition have been studied by high-temperature expansions[70], Monte Carlo simulations[71, 72, 73, 74, 75, 76], Pfaffian or combinatorial methods[77, 78], and numerical transfer-matrix calculations on small systems[69, 79]. The phase transition is signalled by a diverging correlation length ξ as $T \rightarrow 0$; one also finds algebraically decaying correlations between spins in the ground state. A possible experimental realization of the 2D $\pm J$ spin glass ($Rb_2Cu_{1-x}Co_xF_4$) has been studied by Dekker et al[80].

This chapter presents a new algorithm to study the 2D $\pm J$ spin glass. Our calculations, like the earlier ones of Blackman and Poulter[77] and Inoue[78], are based on the combinatorial expansion for 2D Ising models[23]. Unlike these authors, however, we use the combinatorial expansion to compute entire partition functions for spins on a square lattice; in particular, our algorithm returns the density of states as an exact integer result. An important feature of algorithms based on the combinatorial method is that they execute in polynomial time. This distinguishes them from the numerical transfer-matrix algorithm of Morgenstern and Binder[69], which must keep track of 2^L spin configurations in order to compute the partition function on a strip of width L . Our algorithm should also be compared to various integer transfer-matrix algorithms[81, 82, 83, 84] that have appeared in recent years. We obtain exact results on square lattices much too large to be tackled by transfer-matrix techniques. As with all exact methods, our algorithm can serve as a useful check on the performance of specialized replica[74] and multicanonical[75] techniques for Monte Carlo simulation.

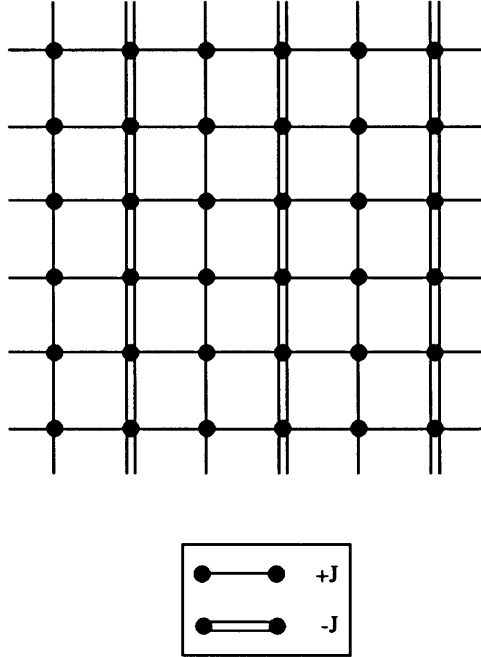


Figure 3.1: Distribution of bonds for the fully frustrated Ising model.

Knowledge of the density of states also enables us to compute new quantities, such as the roots of the partition function in the complex plane[82].

For purposes of comparison, we have also used our algorithm to examine the fully frustrated (FF) Ising model in two dimensions. This model has been solved exactly using standard techniques[85]. Like the $\pm J$ spin glass, it undergoes a phase transition at $T = 0$. The critical properties of this transition, however, are well understood[86, 87]: the spin-spin correlation length diverges exponentially as $\xi_{FF} \sim e^{2J/T}$, and the heat capacity at low temperatures behaves as $C_{FF} \sim T^{-3}e^{-4J/T}$. Figure 3.1 shows one possible bond profile for the fully frustrated Ising model. Note that the product $\Phi = J_{ij}J_{jk}J_{kl}J_{li}/J^4$ around each elementary plaquette on the square lattice is equal to -1 .

The rest of this chapter is organized as follows. In Section 3.2, we give a gen-

eral overview of the algorithm itself. We start by reviewing the high-temperature expansion of the partition function and the procedure for counting closed loops on the square lattice. We then discuss the computer implementation of this method as an exact integer algorithm that outputs the density of states. A number of special features make the algorithm a useful complement to well-established techniques such as Monte Carlo simulation and the numerical transfer matrix method.

In Section 3.3, we present our results on the $\pm J$ spin glass. A variety of issues are explored. First, we compare our estimates of the ground-state energy and entropy with those of previous studies on the $\pm J$ spin glass. The results are found to be in good agreement. Second, we investigate the number of low-level excitations on a square lattice with periodic boundary conditions. For the spin glass, we find that difference in entropy between the ground state and the first excited state grows faster than $\ln N$ but slower than $\ln N^2$, where $N = L^2$ is the number of spins. We examine the consequences of this for the low-temperature behavior of the heat capacity. Third, using the complete density of states, we compute the roots of partition functions in the complex temperature plane. We relate the finite-size scaling of the smallest roots to the divergence of the correlation length as $T \rightarrow 0$. The results suggest an exponentially diverging correlation length $\xi \sim e^{2\beta J}$ in contrast to previous works. Finally, motivated by scaling theories[64, 65, 100, 101] of spin glasses, we examine the properties of defects at $T = 0$ in the $\pm J$ spin glass. The probability of non-zero defect energies on a square lattice of side L is found to scale as $p(L) \sim L^{-\eta}$, with $\eta = 0.22 \pm 0.06$. Likewise, we find that the defect entropy scales as $\delta S_L \sim L^{y_S}$, with $y_S = 0.49 \pm 0.02$.

In Section 3.4, we mention possible extensions of this work to $\pm J$ models with varying levels of frustration and/or missing bonds. We also discuss the potential for polynomial algorithms to investigate two-dimensional Ising models with other types of quenched randomness[78].

The appendices contain discussion of technical points. Appendix A explains the handling of periodic boundary conditions. Appendix B describes various aspects of implementation, including large integer arithmetic, sparse matrix multiplication, power series manipulations, and special symmetries.

3.2 An Exact Integer Algorithm

We consider a system of Ising spins $\sigma_i = \pm 1$ on an $L \times L$ square lattice. The Hamiltonian is given by

$$\mathcal{H} = - \sum_{\langle ij \rangle} J_{ij} \sigma_i \sigma_j, \quad (3.2)$$

where the sum is over all pairs of nearest neighbor spins. The quenched random bonds $\{J_{ij}\}$ are chosen from the bimodal distribution¹

$$P(J_{ij}) = \frac{1}{2} \delta(J_{ij} - J) + \frac{1}{2} \delta(J_{ij} + J), \quad (3.3)$$

¹In practice, we also imposed the global constraint that exactly one-half of the plaquettes on the square lattice were frustrated. For the bimodal distribution in eq. (3.3), the probability for a plaquette to be frustrated is equal to one-half. In the limit of infinite size, the concentration of frustrated plaquettes x_F also tends to this value. The restriction to realizations with $x_F = 1/2$ reduces the statistical error that arises in the computation of quenched averages over the bond disorder[69]. We cannot compute these averages exactly, but must resort to sampling a finite number of realizations of randomness. In practice, one generates the $\pm J$ bonds independently from the distribution in eq. (3.3), then discards configurations that do not meet this criterion.

with $J > 0$. On a lattice with periodic boundary conditions, there are exactly $2N$ bonds, with $N = L^2$ the total number of spins.

The partition function of the system is given by

$$Z = \sum_{\{\sigma_i\}} e^{-\beta\mathcal{H}}, \quad (3.4)$$

with $\beta \equiv 1/T$. A high temperature expansion for the partition function in powers of $\tanh(\beta J)$ is facilitated by the identity[88]

$$e^{\beta J_{ij}\sigma_i\sigma_j} = \cosh(\beta J)[1 + s_{ij}\sigma_i\sigma_j \tanh(\beta J)], \quad (3.5)$$

where $s_{ij} \equiv J_{ij}/J$ is equal to the sign of J_{ij} . Eq. (3.5) makes use of the fact that the product $\sigma_i\sigma_j$ can only assume the values ± 1 . Substituting into the expression for the partition function yields

$$Z = \cosh^{2N}(\beta J) \sum_{\{\sigma_i\}} \prod_{(ij)} [1 + s_{ij}\sigma_i\sigma_j \tanh(\beta J)]. \quad (3.6)$$

We have thus transformed the problem of evaluating the partition function from a sum over Boltzmann weights $e^{-\beta\mathcal{H}}$ to a sum over polynomials in $\tanh(\beta J)$, each of order $2N$.

Expanding the product in eq. (3.6) gives rise to 2^{2N} terms of varying order in $\tanh(\beta J)$. Note that a term of ℓ -th order in $\tanh(\beta J)$ has a coefficient of the form $(s_{i_1j_1}\sigma_{i_1}\sigma_{j_1}) \cdots (s_{i_\ell j_\ell}\sigma_{i_\ell}\sigma_{j_\ell})$. There exists a one-to-one mapping in which each ℓ -th order term in $\tanh(\beta J)$ is associated with the set of ℓ bonds $\{J_{i_1j_1}, J_{i_2j_2}, \dots, J_{i_\ell j_\ell}\}$

that determine (along with the spin variables) the sign of its coefficient. This mapping provides a diagrammatic representation of the terms in the high temperature expansion.

The sum over spin configurations can now be performed term-by-term by repeated application of the rule

$$\sum_{\sigma=\pm 1} \sigma^n = \begin{cases} 2 & \text{if } n \text{ even} \\ 0 & \text{if } n \text{ odd} \end{cases}. \quad (3.7)$$

It is straightforward to show that the only terms which survive this procedure are those whose diagrams on the square lattice can be represented as the union of non-overlapping closed loops. Some examples are shown in Figure 3.2. Each of these diagrams contributes an amount $\pm 2^N \tanh^\ell(\beta J)$ to the partition function. The sign is positive (negative) if the diagram contains an even (odd) number of antiferromagnetic bonds; ℓ is the total number of bonds in the diagram. The final result for the partition function thus takes the form

$$Z = 2^N \cosh^{2N}(\beta J) \sum_{\ell=0}^{2N} A_\ell \tanh^\ell(\beta J), \quad (3.8)$$

where the coefficients A_ℓ are pure integers².

Motivated by the diagrammatic representation, Kac and Ward[23] transformed the problem of summing the high temperature series into one of evaluating a local random walk. In particular, they showed that a $4N \times 4N$ hopping matrix could be

²Note that for even-valued L , closed loops on the square lattice necessarily traverse an even number of bonds; as a result, odd powers of $\tanh(\beta J)$ do not appear in the high-temperature expansion. Because this simplifies the algebra considerably, we only consider lattices with even-valued L .

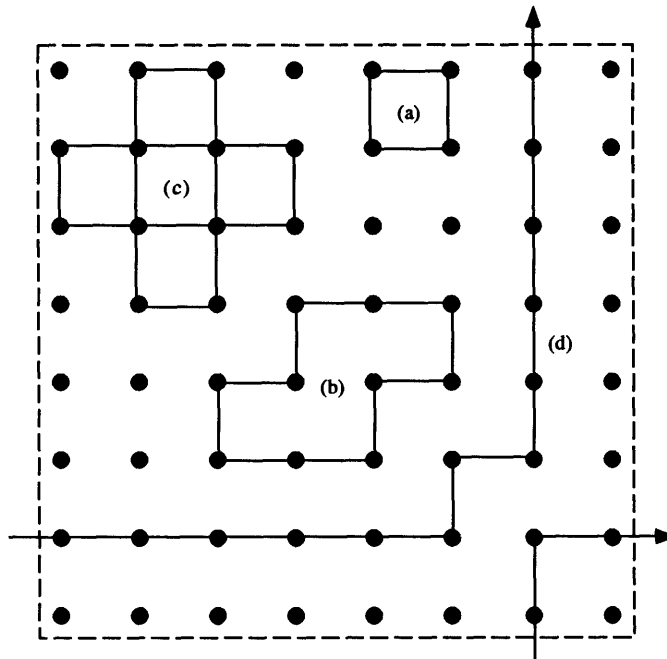


Figure 3.2: Closed graphs of length (a) $\ell = 4$, (b) $\ell = 10$, and (c,d) $\ell = 16$ on the square lattice with periodic boundary conditions. Each diagram represents a term $\tanh^\ell(\beta J)$ in the high-temperature expansion of the partition function. Graphs that wrap around the lattice such as (d) require special handling in performing the sum over diagrams.

used to compute the coefficients of the high-temperature expansion on the square lattice. Though they focused on the problem of the pure 2D Ising model with no randomness, the method they proposed is quite general. It has been used to study periodically frustrated models[85, 89, 90], and more recently, the 2D Ising spin glass[77, 78]. In what follows, we review only the most basic elements of this method. The reader who wishes a more complete understanding should consult any of several earlier treatments[88, 91, 92].

We continue to focus on an $L \times L$ square lattice with periodic boundary conditions and quenched random bonds. The reduced bond variables $s_{ij} = s_{ji} = J_{ij}/J$ are equal to $+1$ for ferromagnetic bonds and -1 for antiferromagnetic bonds. We use the

notation $i\mu$ to refer to the *directed* bond that points in the μ direction and terminates at the i th spin with coordinates (x_i, y_i) . Here, $1 \leq x_i, y_i \leq L$, and $\mu \in \{\uparrow, \leftarrow, \downarrow, \rightarrow\}$.

On a square lattice with $N = L^2$ sites, there are $4N$ directed bonds.

The $4N \times 4N$ Kac-Ward matrix evolves a random walker from one bond to the next, keeping track of any changes in her local orientation. In this way, the matrix generates an amplitude for each random walk that can be related to the weight of the corresponding diagram in the high-temperature series. To avoid overcounting, the amplitude of each random walk must acquire a factor of -1 at each self-intersection. On the square lattice, this is done by employing Whitney's theorem[93] and assigning a factor $e^{i\pi/4}$ ($e^{-i\pi/4}$) to each clockwise (counterclockwise) turn of the walk[88, 91]. The matrix elements $U_{j\sigma, i\mu}$ are shown in Table 3.1. The final result of Kac and Ward, valid in the thermodynamic limit of $N \rightarrow \infty$, is that

$$\frac{\ln Z}{N} = \ln[2 \cosh^2(\beta J)] - \frac{1}{2N} \sum_{\ell=0}^{\infty} \frac{1}{\ell} \text{tr}(U^\ell) \tanh^\ell(\beta J), \quad (3.9)$$

where $\text{tr}(U^\ell)$ denotes the trace of the ℓ th power of the hopping matrix. These traces count the closed loops on the square lattice in such a way as to reproduce the high temperature expansion of the partition function. Eq. (3.9) can be written in the more compact form

$$Z = 2^N \cosh^{2N}(\beta J) \sqrt{\det[1 - U \tanh(\beta J)]}, \quad (3.10)$$

using standard matrix identities. The problem is thus reduced to evaluating the determinant of a $4N \times 4N$ matrix. For models with 2D translational symmetry (e.g.

μ	σ	$s_{ij}U_{j\sigma,i\mu}$
\uparrow	\uparrow	$\delta_{x_i,x_j}\delta_{y_i+1,y_j}$
\rightarrow	\rightarrow	$\delta_{x_i+1,x_j}\delta_{y_i,y_j}$
\downarrow	\downarrow	$\delta_{x_i,x_j}\delta_{y_i-1,y_j}$
\leftarrow	\leftarrow	$\delta_{x_i-1,x_j}\delta_{y_i,y_j}$
\uparrow	\rightarrow	$\delta_{x_i+1,x_j}\delta_{y_i,y_j}e^{i\pi/4}$
\rightarrow	\downarrow	$\delta_{x_i,x_j}\delta_{y_i-1,y_j}e^{i\pi/4}$
\downarrow	\leftarrow	$\delta_{x_i-1,x_j}\delta_{y_i,y_j}e^{i\pi/4}$
\leftarrow	\uparrow	$\delta_{x_i,x_j}\delta_{y_i+1,y_j}e^{i\pi/4}$
\uparrow	\leftarrow	$\delta_{x_i-1,x_j}\delta_{y_i,y_j}e^{-i\pi/4}$
\leftarrow	\downarrow	$\delta_{x_i,x_j}\delta_{y_i-1,y_j}e^{-i\pi/4}$
\downarrow	\rightarrow	$\delta_{x_i+1,x_j}\delta_{y_i,y_j}e^{-i\pi/4}$
\rightarrow	\uparrow	$\delta_{x_i,x_j}\delta_{y_i+1,y_j}e^{-i\pi/4}$
\uparrow	\downarrow	0
\downarrow	\uparrow	0
\leftarrow	\rightarrow	0
\rightarrow	\leftarrow	0

Table 3.1: Elements, $U_{j\sigma,i\mu}$, of the $4N \times 4N$ Kac-Ward hopping matrix for the 2D $\pm J$ spin glass. The reduced bond variables s_{ij} are equal to +1 for ferromagnetic bonds and -1 for antiferromagnetic bonds.

the pure Ising model with no randomness or the fully frustrated model), one can use a plane-wave basis to diagonalize the matrix and obtain analytic results in the thermodynamic limit. Of course, the plane-wave basis does not diagonalize the matrix for a system with quenched random bonds, such as the $\pm J$ spin glass.

We can, however, use the Kac-Ward method to compute the partition function for a $\pm J$ spin glass of *finite size*[77]. In this case, Eq. (3.10) is no longer valid, but must be slightly modified to incorporate the choice of periodic boundary conditions. As it stands, Eq. (3.10) does not properly weight diagrams which cross the boundary of the lattice. The correct result, based on the combinatorics of closed loops on periodic

lattices[94], is $Z = (-Z_1 + Z_2 + Z_3 + Z_4)/2$ with

$$Z_\lambda = 2^N \cosh^{2N}(\beta J) \sqrt{\det[1 - U_\lambda \tanh(\beta J)]}. \quad (3.11)$$

Here, U_λ are four distinct $4N \times 4N$ hopping matrices related to one another by boundary transformations (see Appendix A). The linear combination $(-Z_1 + Z_2 + Z_3 + Z_4)/2$ ensures that all diagrams, including those which loop around the entire lattice, are weighted correctly in the final expression for the partition function. Further details on handling periodic boundary conditions are discussed in Appendix A.

We have implemented this algorithm on the computer as follows. Given a set of bonds, $\{J_{ij}\}$, we first construct the $4N \times 4N$ matrices U_λ and compute the traces $\text{tr}[(U_\lambda)^\ell]$ for $\ell \leq N$. This step of the algorithm is the most computationally intensive. The coefficients of the series expansions for $\ln Z_\lambda$ are related to the traces by eq. (3.9). Next, we compute the high temperature series for Z . This is done by exponentiating the series for $\ln Z_\lambda$ and taking the linear combination that incorporates periodic boundary conditions. The high temperature expansion for Z is a polynomial in $\tanh(\beta J)$ with integer coefficients; the last term, of order $2N$, is derived from the graph that traverses every bond on the square lattice. These $2N$ coefficients have an end-to-end symmetry (see Appendix B) that enables one to compute them from the first N powers of the hopping matrix. Finally, we expand powers of $\cosh(\beta J)$ and $\tanh(\beta J)$ and rewrite Z as a polynomial in $e^{-\beta J}$; the end result $Z = \sum_E g(E) e^{-\beta E}$ yields the density of states. For an Ising model with $\pm J$ bonds, we can perform all these operations using only integer arithmetic. Each of these steps is discussed

further in Appendix B.

The algorithm has several desirable features. First, it returns the partition function Z as an exact integer result. In this way, it not only avoids the statistical uncertainties inherent in Monte Carlo simulation; it also avoids the floating point errors that creep into numerically “exact” calculations of Z in large systems at low temperatures. The algorithm thus provides us with an efficient and reliable way to investigate energies and entropies at low temperatures. This is particularly important in a system that exhibits a phase transition at $T = 0$, such as the $\pm J$ spin glass. We can also calculate other quantities, such as the roots of the partition function in the complex plane, or the number of low-level excitations, that are otherwise inaccessible. Unfortunately, the necessity of handling large integers (of order 2^N) complicates what would otherwise be a rather straightforward algorithm to implement. We gratefully acknowledge the existence of two software tools that proved helpful in this regard, the BigNum package of C routines and the commercial program MATHEMATICA[95].

A second advantage of the algorithm is that it executes in polynomial time. This stands in contrast to the numerical transfer matrix method of Morgenstern and Binder[69], which has time and memory requirements that grow exponentially with system size. Computing the powers of the hopping matrix requires $O(N^3)$ arithmetic operations (see Appendix B), while in general, the power series manipulations take much less time. The computation time of our method scales slightly worse than N^3 because the arithmetic operations involve integers of order 2^N . Roughly speaking, the processing time for storing and manipulating these integers grows linearly in N , that is, with the number of digits necessary to represent them in memory. One might

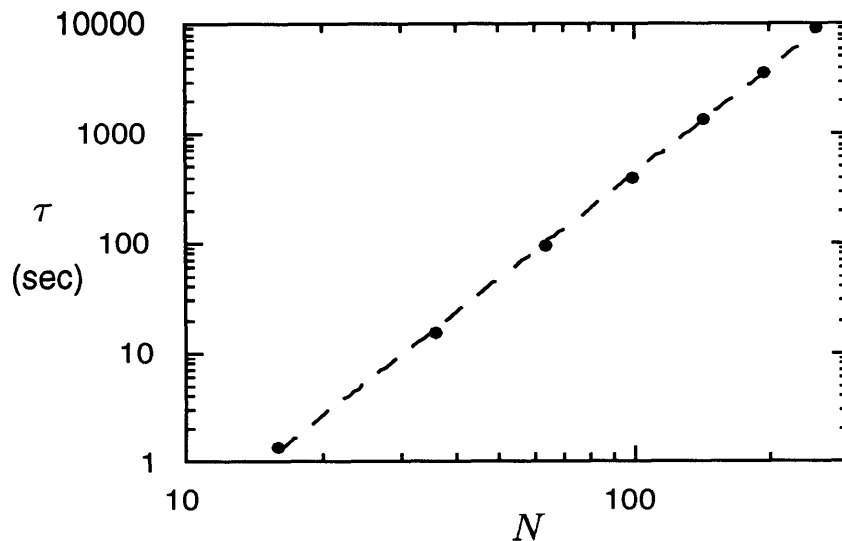


Figure 3.3: Log-log plot of computation time versus number of spins on the square lattice. The fit shows $\tau \sim N^{3.17}$.

therefore expect the computation time to scale as $\tau \sim N^{3+\delta}$, with $0 < \delta < 1$. Some sample times for lattices of various size are shown in Figure 3.3; we find an effective value of $\delta \approx 0.17$. We obtained most of our results on dedicated DEC 3100 workstations; the largest lattice that we examined had $N = 36 \times 36$ spins. Perhaps most intriguing is that the computation of the traces can be trivially decomposed into $O(N)$ independent computations, so that a faster, parallel implementation of the algorithm (on a supercomputer or spread across several workstations) should be possible.

3.3 Results

We examined the $\pm J$ spin glass on lattices of size $L = 4$ to $L = 36$. The number of realizations of randomness studied for each lattice size is shown in Table 3.2.

We performed quenched averages by assigning an equal probability to each random sample:

$$\bar{\theta} = \frac{1}{S} \sum_{s=1}^S \theta_s. \quad (3.12)$$

Our results on the $\pm J$ spin glass show that our algorithm is a useful complement to well-established techniques such as the numerical transfer-matrix method[69] and Monte Carlo simulation[74, 75]. We also present some results for the 2D fully frustrated Ising model[85, 86]. The comparison of these two systems, both of which undergo phase transitions at $T = 0$, is often revealing.

The typical output of our algorithm is shown in Table 3.3, which gives the density of states for a 10×10 system with random $\pm J$ bonds. Knowledge of the density of states enables us to perform calculations in either the microcanonical or canonical ensemble. In the limit of infinite size, of course, the two ensembles should yield identical results. Figure 3.4 shows the entropy versus energy curves for the system using both methods. The difference between the two calculations provides a sense of the magnitude of finite-size effects at $L = 10$. The heat capacity of the system, calculated in the canonical ensemble, is shown in Figure 3.5.

lattice size	samples
4, 6, 8	8000
10, 12, 14	2000
16, 18	800
20, 22, 24	80
32, 36	4

Table 3.2: Number of realizations of randomness studied for each lattice size.

E	$g(E) = g(-E)$
-142	1714
-138	393704
-134	26810830
-130	1081037598
-126	30884856124
-122	681219831812
-118	12208593158206
-114	183664270660180
-110	2370651664443342
-106	26665546247842738
-102	264439620156521442
-98	2332975127519659794
-94	18440896548918449182
-90	131332616729581167390
-86	846428717389978116910
-82	4953662247974988779202
-78	26396299869918456764966
-74	128336521283648955046794
-70	570254120321897569640274
-66	2318884696771718964963738
-62	8639021129076421274926654
-58	29514583986932577806279238
-54	92546253453945502148308030
-50	266538596843504319655793666
-46	705577194982709216381878662
-42	1717894672802780124241974506
-38	3849313192834817644425385210
-34	7942459215428479616883558298
-30	15098868015221922648819834550
-26	26458472739977677505478377606
-22	42757102380960530199403590582
-18	63744656075332650586700499562
-14	87703140136204818329708230226
-10	111388263537730445390041718418
-6	130618789608427927645846927382
-2	141440475064667109660456174158

Table 3.3: Energies, E , and numbers of states, $g(E)$, for a $\pm J$ spin glass on a 10×10 lattice with periodic boundary conditions. The calculation took 110 seconds on an INDIGO 4000 workstation with no parallelization.

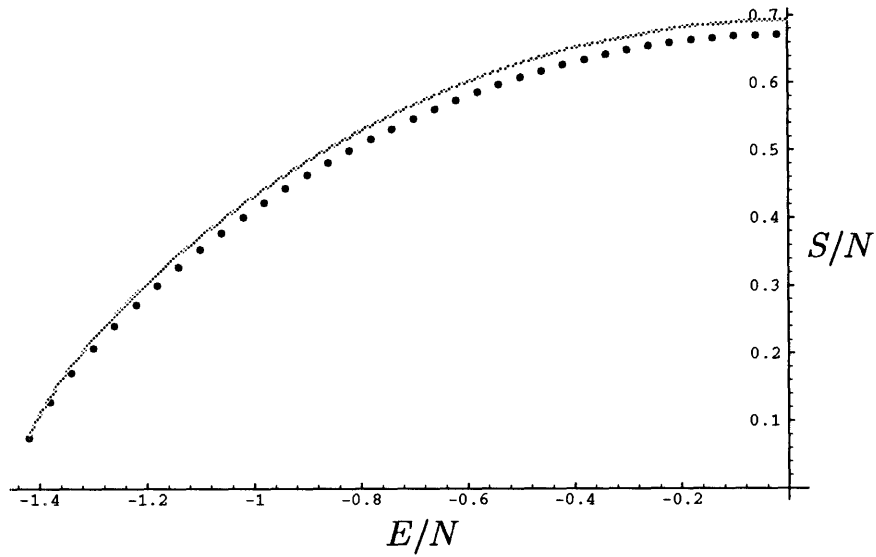


Figure 3.4: Energy versus entropy for a $10 \times 10 \pm J$ spin glass. The energy levels in the microcanonical ensemble occur at discrete intervals of $4J$. The smooth curve is the result from the canonical ensemble.

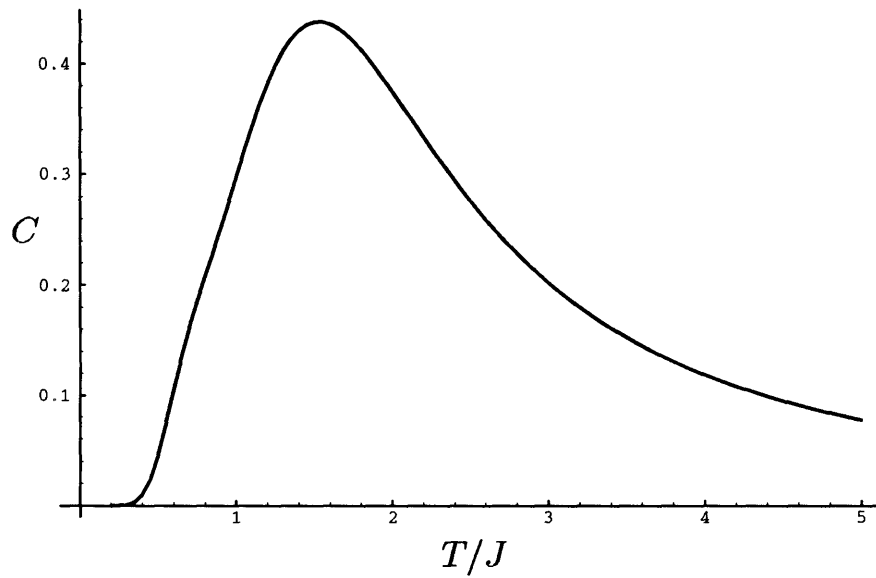


Figure 3.5: Heat capacity versus temperature for a $10 \times 10 \pm J$ spin glass.

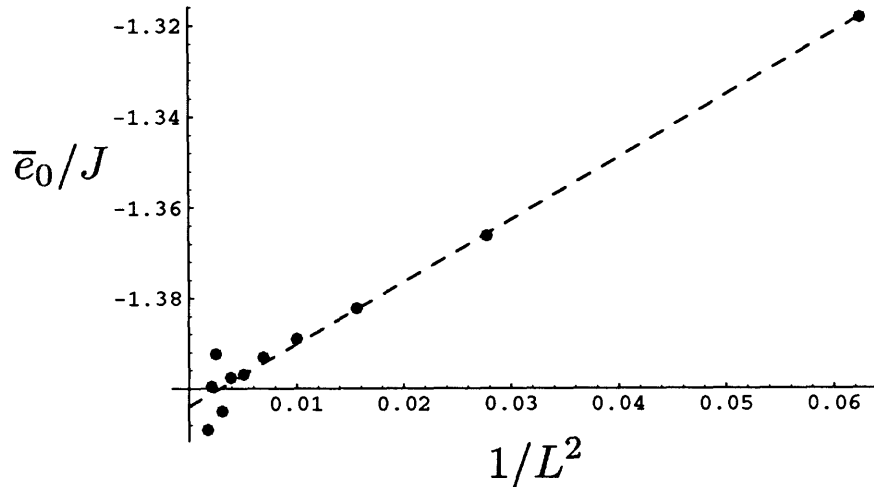


Figure 3.6: Average ground-state energy density versus $1/L^2$ for the $\pm J$ spin glass.

3.3.1 Ground States and Low-Level Excitations

We used the algorithm first to study the thermodynamic properties of the ground state. We fit the data for the ground-state energy and entropy to the finite-size form $f_L = f_\infty + a/L^2$; a linear term in $1/L$ is not expected due to the periodic boundary conditions. The ground state energy per spin \bar{e}_0 is plotted versus $1/L^2$ in Figure 3.6. Extrapolating to infinite lattice size, we estimate that $\bar{e}_0/J = -1.404 \pm 0.002$. Our value compares favorably with previous results from combinatorial techniques[77, 78], numerical transfer-matrix calculations[79], and Monte Carlo simulations[74, 75, 76]. The most accurate of these[77] reports $\bar{e}_0/J = -1.4020 \pm 0.0004$. Likewise, we plot the ground state entropy per spin in Figure 3.7. Here, we find $\bar{s}_0 = 0.075 \pm 0.002$, in fair agreement with the best previous estimate[77], $\bar{s}_0 = 0.0704 \pm 0.0002$. The discrepancy may be due to a number of factors, including finite-size effects, boundary conditions, and insufficient quench-averaging.

We also used the algorithm to study the number of low-level excitations in the

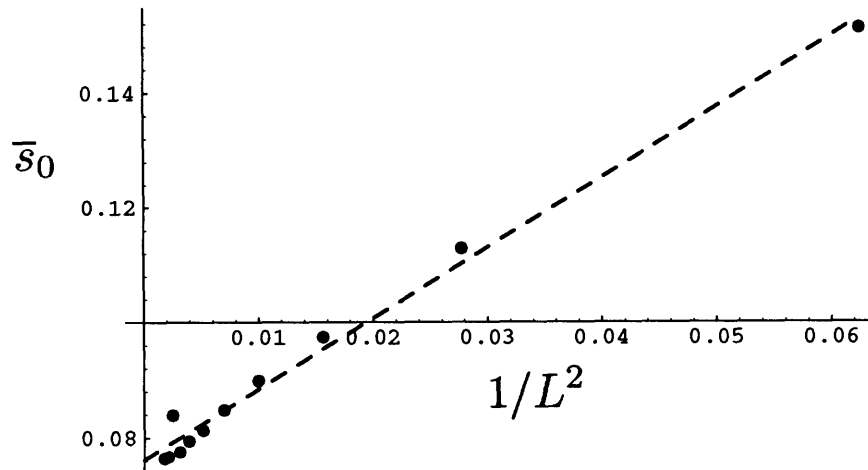


Figure 3.7: Average ground-state entropy density versus $1/L^2$ for the $\pm J$ spin glass.

$\pm J$ spin glass. On a lattice with periodic BCs, the lowest excited state has an energy $4J$ above the ground state. The quantity $e^{\Delta S} = g(E_0 + 4J)/g(E_0)$ measures the degeneracy ratio of excited and ground states. Figure 3.8 shows a semilog plot of $\overline{\Delta S}_{SG}$ versus the number of spins N . The fact that $\overline{\Delta S}_{SG}$ grows faster than $\ln N$ (dashed line) suggests that the low-lying excitations of the $\pm J$ spin glass involve spin flips on large length scales.

The abundance of low-lying excitations affects the low-temperature behavior of the heat capacity. In a finite system with energy gap $4J$, the heat capacity vanishes as $C \sim \beta^2 e^{-4\beta J}$. As pointed out by Wang and Swendsen[74], this behavior can break down in the thermodynamic limit. The 1D Ising model with periodic BCs shows how this can happen: the energy gap is $4J$, but the heat capacity of an infinite system vanishes as $C_{1D} \sim \beta^2 e^{-2\beta J}$. The anomalous exponent reflects the fact that the number of lowest excited states grows as N^2 . From MC and TM studies, Wang and Swendsen[74] conclude that $\overline{C}_{SG} \sim \beta^2 e^{-2\beta J}$ for the 2D $\pm J$ spin glass as well. For

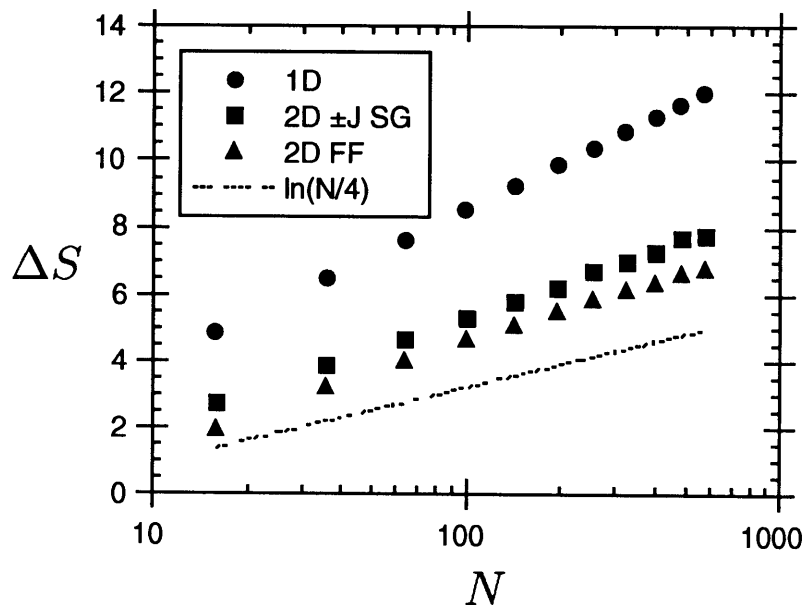


Figure 3.8: Finite-size scaling of the entropy difference between the ground state and the lowest excited states for the 1D Ising model, the 2D FF model, and the 2D $\pm J$ spin glass (quench-averaged).

purposes of comparison, we have included data for the 1D Ising model in Figure 3.8.

The plots show that $\overline{\Delta S}_{SG}$ grows faster than $\ln N$ but slower than $\ln N^2$.

The disagreement in slope between ΔS_{1D} and $\overline{\Delta S}_{SG}$ leads us to suggest a different form for \overline{C}_{SG} . As motivation, we appeal to another exactly soluble model with a phase transition at $T = 0$: the fully frustrated (FF) Ising model on a square lattice[85]. On a periodic lattice, the lowest excited states of the FF model have energy $4J$ above the ground state. Low-lying excitations above the ground state in the FF model also involve spin flips on large length scales: Figure 3.9 shows that the ratio $g(E_0 + 4J)/g(E_0)$ grows as $N \ln N$. The large number of low-lying excitations causes the heat capacity to vanish as $C_{FF} \sim \beta^3 e^{-4\beta J}$. Note the extra power of temperature.

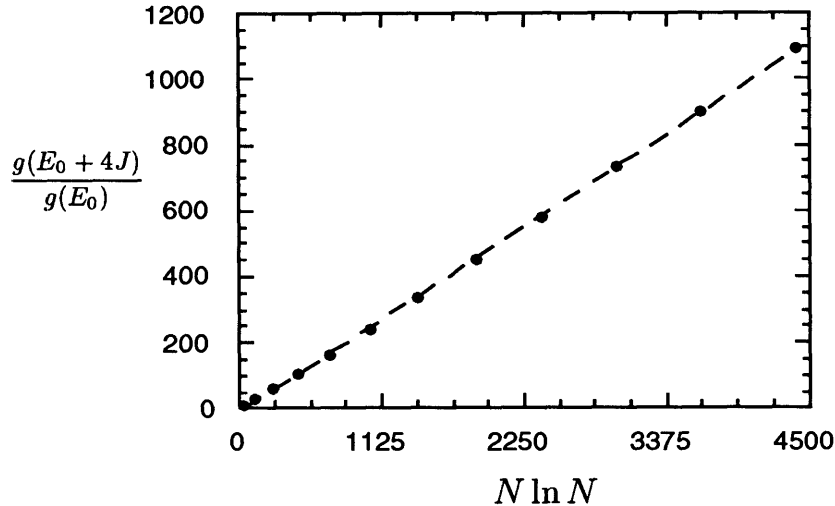


Figure 3.9: Ratio $g(E_0 + 4J)/g(E_0)$ versus $N \ln N$ for the fully frustrated model.

Comparing ΔS_{FF} and $\overline{\Delta S}_{SG}$ in Figure 3.8, we suspect a similar behavior may describe the $\pm J$ spin glass, so that $\overline{C}_{SG} \sim \beta^{2+\rho} e^{-4\beta J}$ with $\rho > 0$.

To explore this possibility, we calculated quench-averaged heat capacities in the canonical ensemble, using MATHEMATICA[95] to do the high-precision floating point arithmetic. A semilog plot of the reduced heat capacity $(\beta J)^2 \overline{C}_{SG}$ versus temperature is shown in Figure 3.10 for lattices of size $L = 12, 16,$ and 20 . In this plot, the slope of the data measures the value a in the exponent $e^{a\beta J}$ of the low-temperature heat capacity. Wang and Swendsen reported that $a \rightarrow -2$ at low-temperatures, based on Monte Carlo data from 4-17 samples of $L = 64$ and $L = 128$ lattices. Unlike Wang and Swendsen, we do not find that this slope bends over to a value of -2 at low temperatures. Rather, our data for the spin glass, like that for the fully frustrated model, suggests the exponential decay $C \sim e^{-4\beta J}$. The discrepancies between the two sets of data could be due a number of factors: finite-size effects, unreliable Monte Carlo

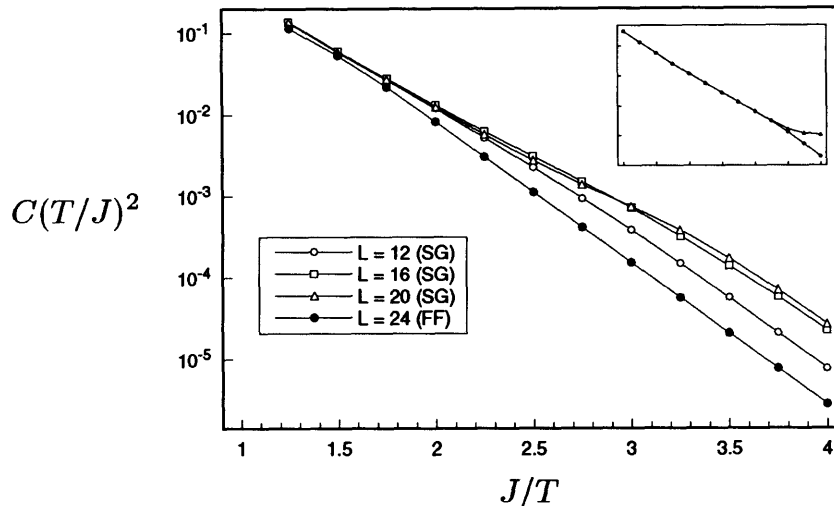


Figure 3.10: Semilog plot of the reduced heat capacity $C(T/J)^2$ versus J/T for the FF model and the spin glass (quench-averaged). The corner plot shows two calculations for the $L = 24$ spin glass; the lower curve was computed after deleting just one sample from the quench-average.

estimates, and/or insufficient quench-averaging. The last point deserves emphasis. At low temperatures, the heat capacity is determined by the ratio $g(E_0 + 4J)/g(E_0)$, a quantity which exhibits very large fluctuations between different realizations of randomness. To illustrate this, we have shown in the upper corner of Figure 3.10 two calculations of \overline{C}_{SG} for $L = 24$, one averaged over seventy-nine realizations of randomness, the other over eighty. Note how one realization of randomness dominates the sample average of the heat capacity at low temperatures. One expects the sample-to-sample fluctuations to increase in magnitude with system size. On larger lattices, the effects of small sample sizes could be quite severe.

Another probe of the low-temperature heat capacity is the number of excitations on a lattice with *free* boundary conditions. On such a lattice, the spin glass has excitations of energy $2J$ above the ground state. Wang and Swendsen used the nu-

merical transfer-matrix method[69] to calculate the ratio $N^{-1}\overline{g(E_0 + 2J)/g(E_0)}$ on square lattices with free boundary conditions. One expects a constant value for this ratio if $\overline{C}_{SG} \sim \beta^2 e^{-2\beta J}$, a slow decay with system size if $\overline{C}_{SG} \sim \beta^{2+\rho} e^{-4\beta J}$. Wang and Swendsen observed that this ratio varied weakly with system size and concluded that the low-temperature heat capacity in the thermodynamic limit behaves as $\overline{C}_{SG} \sim \beta^2 e^{-2\beta J}$. Figure 3.11 shows a plot of $N^{-1}\overline{g(E_0 + 2J)/g(E_0)}$ versus lattice size for the $2d \pm J$ spin glass with free boundary conditions. Our sample average was performed over 200 realizations of randomness for lattices of size $L = 4$ to 14. Also included on the same plot are Wang and Swendsen's results and the corresponding data for the fully frustrated model. In the FF model, the ratio of $2J$ excitations to ground states seems to decay as $L^{-1} \ln L$. Note that the ratio for the spin glass exhibits an irregular size dependence; this may be due to insufficient quench-averaging and/or finite-size effects. Given the smallness of the lattice sizes, we do not feel that either set of data distinguishes between a constant value or a slow asymptotic decay, such as $L^{-1}(\ln L)^\rho$.

The situation for the spin glass heat capacity \overline{C}_{SG} can therefore be summarized as follows: The large number of $4J$ excitations leads to an anomalous behavior at low temperatures. A similar effect is observed in the 1D Ising model, where $C_{1D} \sim \beta^2 e^{-2\beta J}$, and the 2D FF model, where $C_{FF} \sim \beta^3 e^{-4\beta J}$. For the spin glass, one can either calculate the heat capacity directly, or look at the number of excitations on lattices with open boundary conditions. The available evidence does not clearly distinguish between $e^{-2\beta J}$ and $e^{-4\beta J}$ behavior, but in our view, the analogy to the fully frustrated case is more compelling. It suggests that at low temperatures, the

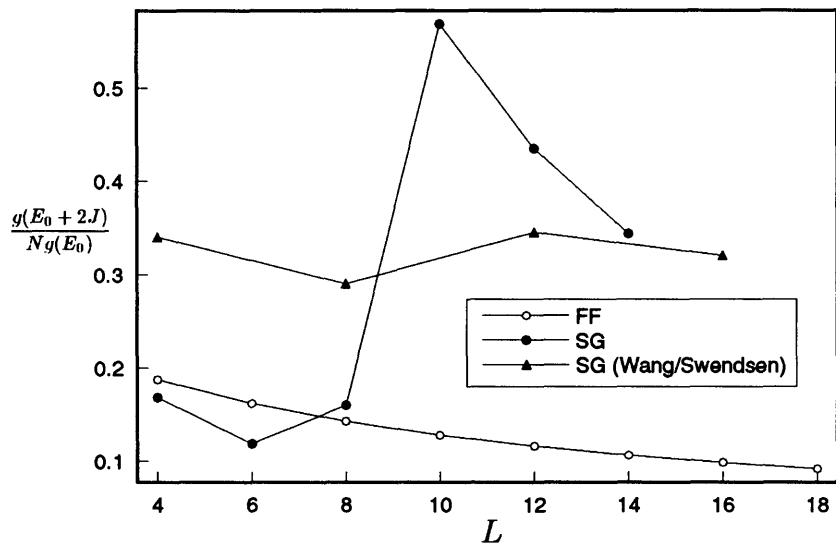


Figure 3.11: Normalized ratio of $2J$ excitations to ground states versus lattice size for the FF model and spin glass with open boundary conditions.

heat capacity for the 2D $\pm J$ spin glass behaves as $\overline{C}_{SG} \sim \beta^{2+\rho} e^{-4\beta J}$, with $\rho > 0$.

3.3.2 Roots of Z in the Complex Plane

One way to investigate phase transitions is to examine the roots of the partition function, Z , in the complex plane. This was first done by Fisher[96] in a study of the 2D Ising model with uniform $+J$ bonds. Fisher calculated the distribution of roots of the partition function in the complex $z = e^{-2\beta J}$ plane. In the limit of infinite lattice size, he showed that the roots condense onto two circles centered at $z = \pm 1$ and related the singular behavior in the free energy to the distribution of roots in the vicinity of the positive real axis. Since a system of finite size does not exhibit non-analytic behavior, it is clear that the roots of the partition function can only condense onto the positive real axis in the thermodynamic limit. An interesting property of the 2D Ising model is that the roots of the component partition functions Z_λ (see

eq. (3.11)) lie on Fisher's loci even if the lattice size is finite [97, 98]. This property is illustrated in Figure 3.12 which shows the roots of the partition functions Z_λ in the complex plane for the 2D Ising model with uniform bonds. Note that roots of Z_λ can lie on the positive real axis, even for lattices of finite size; only in the thermodynamic limit, however, will roots of the actual partition function cross this axis, giving rise to singular behavior in the free energy.

We examined the roots of the component partition functions Z_λ for the $\pm J$ spin glass on lattices of size $L = 4$ to $L = 10$. Figure 3.13 shows some of these roots in the complex $z = e^{-2\beta J}$ plane. (The roots on the positive temperature axis are artifacts of the component partition functions Z_λ .) Note that in considering the ensemble of all realizations, the roots are described by a probability distribution in the complex z plane. Figure 3.13, constructed from a very limited number of realizations, is only intended to convey the most basic features of this probability distribution. For purposes of comparison, we also examined the location of roots in the complex plane for the fully frustrated Ising model. These results are shown in Figure 3.14. In both cases, the roots condense around the origin, indicating a phase transition at zero temperature.

We also examined the roots of the true partition functions for the spin glass; recall that the true partition functions are related to the component partition functions by $Z = (-Z_1 + Z_2 + Z_3 + Z_4)/2$. The zeros of partition functions close to a critical point are subject to finite-size scaling [106]. At a finite-temperature phase transition, such as occurs in the $D \geq 2$ Ising model, the complex zero, $T(L)$, closest to the positive real axis obeys $|T(L) - T_c| \sim L^{-\nu}$; likewise, the correlation length diverges

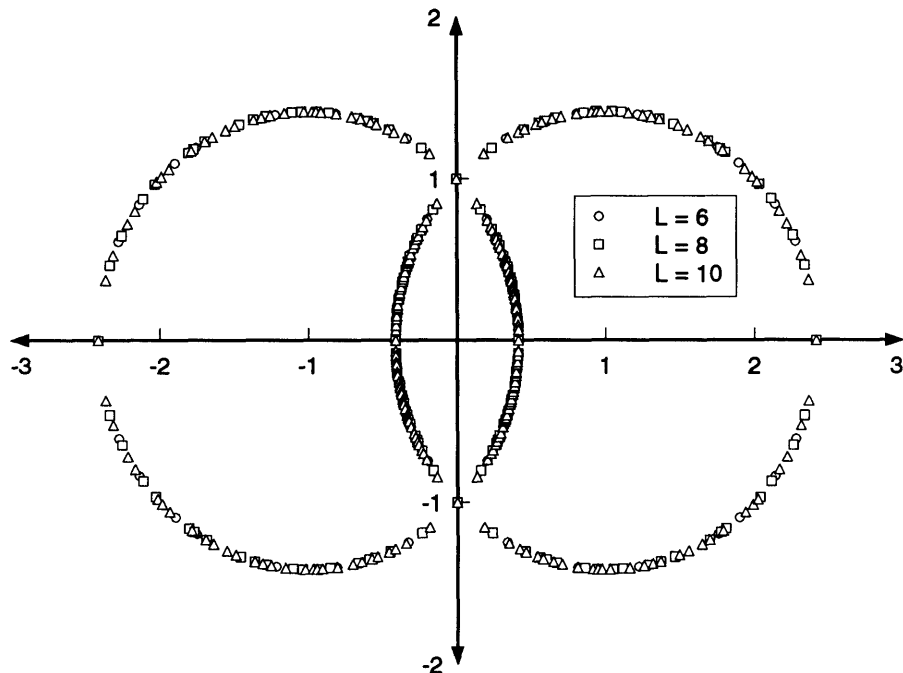


Figure 3.12: Roots of the component partition functions Z_λ in the complex $z = e^{-2\beta J}$ plane for the pure Ising model.

as $\xi \sim (T - T_c)^{-\nu}$, with $\nu = 1/y_t$. (For a study of zeros in the 3D Ising model, see [107].) On the other hand, at a $T = 0$ phase transition, such as occurs in the 1D Ising model, one finds $|z(L)| \sim L^{-y_z}$ with $\xi \sim z^{-1/y_z}$, where $z(L)$ is the smallest root in the complex $z = e^{-2\beta J}$ plane. For example, the partition function for a 1D Ising model with periodic boundary conditions is given by

$$Z = \sum_{\sigma} \exp\{\beta J \sum_i \sigma_i \sigma_{i+1}\} \quad (3.13)$$

$$= 2^L \cosh^L(\beta J) [1 + \tanh^L(\beta J)]. \quad (3.14)$$

The smallest root of Z is $\tanh(\beta J) = e^{\pm i\pi/L}$, or $z(L) = \pm i \tan(\pi/2L)$. As $L \rightarrow \infty$, the magnitude of the root scales as $|z(L)| \sim L^{-1}$, consistent with the fact that the

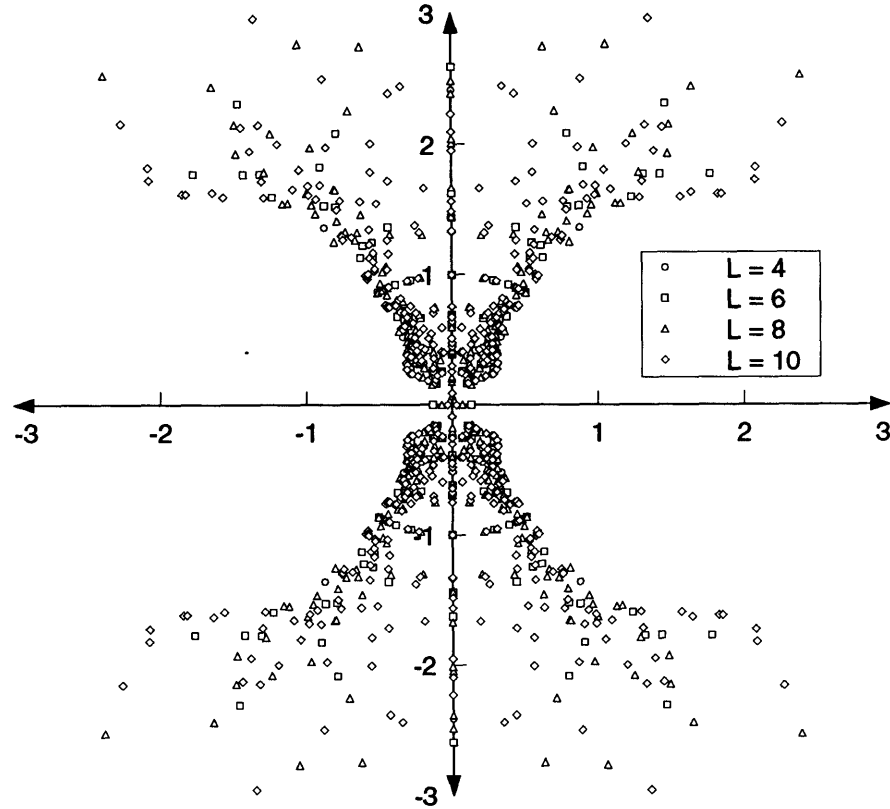


Figure 3.13: Roots of the component partition functions Z_λ in the complex $z = e^{-2\beta J}$ plane for four different realizations ($L = 4, 6, 8,$ and 10) of the $\pm J$ spin glass.

correlation length in the 1D Ising model diverges as $\xi_{1D} \sim e^{2\beta J}$. Similar finite-size scaling is also found in the 2D fully frustrated Ising model. Figure 3.15 shows a log-log plot of $|z(L)|$ versus lattice size for the fully frustrated model; note that the data in this figure correspond to the roots of the true partition functions Z , not those of Z_λ . The magnitude of the smallest root in the complex plane scales as $|z(L)| \sim L^{-1}$, consistent with the exact result $\xi_{FF} \sim e^{2\beta J}$.

In the 2D $\pm J$ spin glass, we observed that, for most realizations of randomness, the smallest root $z(L)$ fell on the imaginary axis. One might expect that the probability distribution for the magnitude of this root would assume a scale-invariant form as

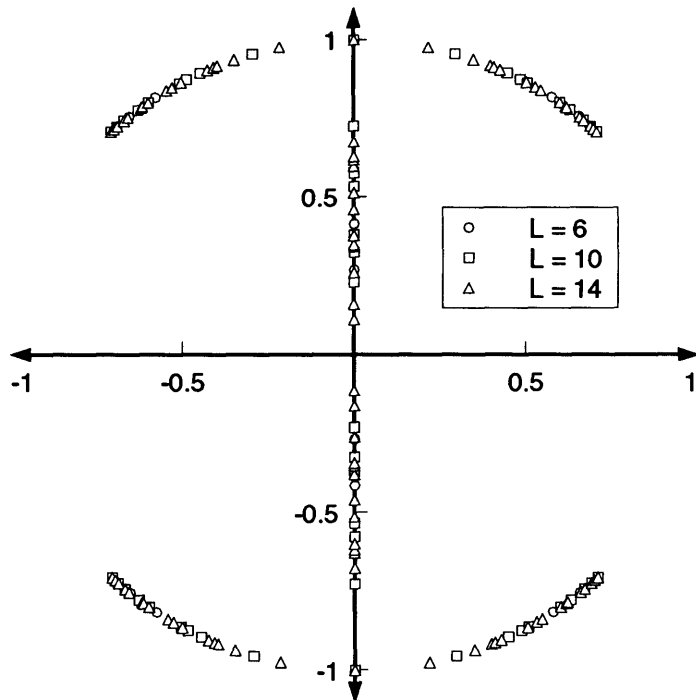


Figure 3.14: Roots of the component partition functions Z_λ in the complex $z = e^{-2\beta J}$ plane for the 2D fully frustrated model.

$L \rightarrow \infty$. We were unable to verify this hypothesis, due to insufficient data on large lattices. Instead, we examined the statistics of $u(L)$, where $u = z^2 = e^{-4\beta J}$. On a square lattice with periodic boundary conditions, the partition function for a 2D $\pm J$ spin glass is polynomial in $e^{-4\beta J}$. We therefore looked at the scaling of roots in the complex $u = e^{-4\beta J}$ plane. Figure 3.16 shows a log-log plot of $|\overline{u(L)}|$ versus lattice size L , where the average was for computational reasons³ restricted to realizations with

³For small systems ($L \leq 12$), we used the commercial program MATHEMATICA[95] to compute the roots of $Z(u) = 0$. For larger systems, this was not feasible. Instead, we used an indirect method to extract the magnitude of the smallest root. Our method works when the desired root has no imaginary component in the complex $u = \exp(-4\beta J)$ plane. This was found to be the case for the majority of partition functions. First, we write the partition function as a polynomial in $u = \exp(-4\beta J)$ with integer coefficients: $Z = \sum_n g_n u^n$. Next, using integer operations, we invert Z and obtain a power series for $1/Z = \sum_m h_m u^m$. The radius of convergence for this series is determined by the smallest root of Z in the complex $u = \exp(-4\beta J)$ plane. The ratio $\lim_{m \rightarrow \infty} |h_m/h_{m+1}|$ is expected to yield the magnitude of this root. If the root is purely real, the

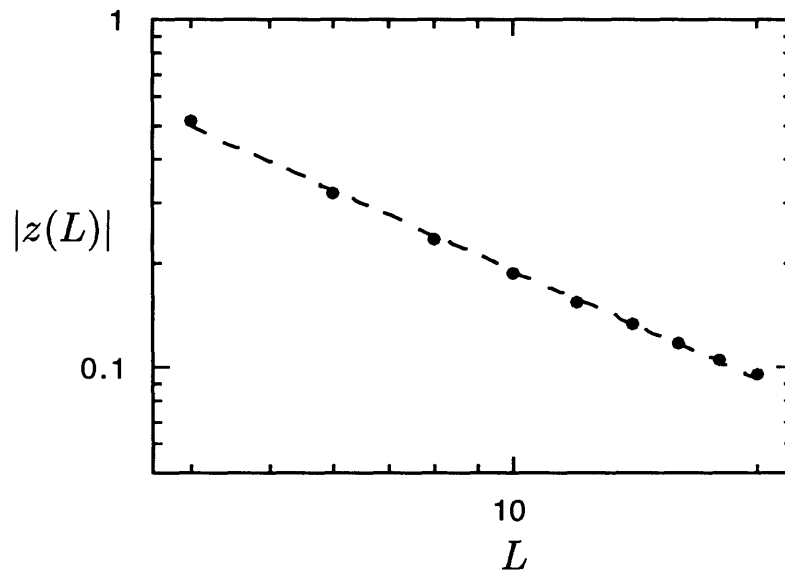


Figure 3.15: Magnitude of the smallest root in the complex $z = e^{-2\beta J}$ plane versus lattice size for the 2D fully frustrated Ising model. The dashed line shows $|z(L)| \sim L^{-1.04}$.

$\text{Im}[u(L)] = 0$ (i.e. with a negative $u(L)$). As in Figure 3.15, the data in this figure correspond to the roots of Z not Z_λ . The fit shows $y_u = 2.2 \pm 0.1$; this suggests to us that the correlation length in the $\pm J$ spin glass diverges as $\xi \sim e^{2\beta J}$. Additional powers of temperature and/or finite-size effects might explain the deviation from $y_u = 2$ in Figure 3.16. Note that this behavior for the correlation length is consistent with hyperscaling, and our claim that, up to powers of temperature, the heat capacity diverges as $C \sim e^{-4\beta J}$. Our result disagrees with previous studies[74, 79, 101] that report $\xi \sim T^{-\nu}$, with $\nu \approx 2.6$ – 2.8 . There may, however, be ambiguities in identifying the correlation length solely on the basis of zeros of $\ln Z$. From comparison to uniform spin models, it is tempting to conclude a critical dimension of $d_l = 2$ from the

ratio converges quite rapidly, making this method an effective one.

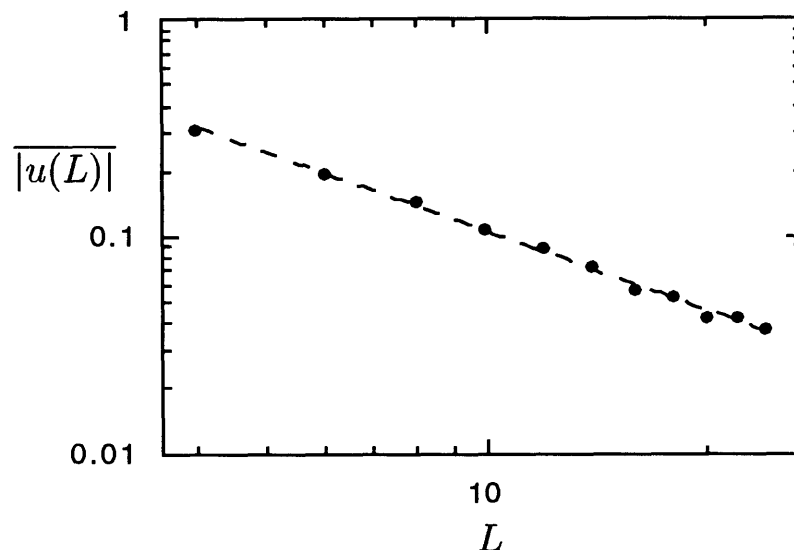


Figure 3.16: Magnitude of the quench-averaged smallest root in the complex $u = e^{-4\beta J}$ plane versus lattice size for the $\pm J$ spin glass. The dashed line shows $|u(L)| \sim L^{-2.2}$.

exponential divergence of the correlation length ξ . There is, however, no compelling theoretical basis for such a conclusion. Indeed, the most recent numerical simulations of the $\pm J$ model in three dimensions[108] do not find evidence for a finite-temperature phase transition⁴.

3.3.3 Defects

A great deal of information on spin glasses has been obtained by examining the defect energies of finite systems at $T = 0$. The defect energy is equal to the difference in free energies of systems with periodic and antiperiodic boundary conditions. At $T = 0$, this reduces to the difference in energy of the ground state. The scaling

⁴See [82], however, for an opposite view based on the behavior of the roots.

theory of defect energies has been developed by Bray and Moore[64, 100], McMillan [101], and Fisher and Huse[65]. One makes the basic ansatz that the defect energy on a $L \times L$ lattice measures the effective coupling between blocks of spins of size L . The variation of the defect energy with the system size then describes the rescaling of the block coupling, or equivalently the temperature, under a transformation of the renormalization group. In a spin glass, the change in boundary conditions is imposed by reversing the signs of L bonds along a single vertical or horizontal strip of the lattice. Bray and Moore[100] used this scaling theory to investigate the 2D Gaussian spin glass on lattices with $L = 2$ to $L = 12$. They found that the probability distribution for the defect energies scaled as

$$P_L(\delta E) = L^{-y_E} f_E\left(\frac{\delta E}{L^{y_E}}\right), \quad (3.15)$$

with $f_E(x)$ a universal function. For the 2D Gaussian spin glass, the exponent y_E is negative, indicating a zero-temperature phase transition with critical exponent $\nu = 1/|y_E|$. Here, ν is the exponent that describes the divergence of the correlation length $\xi \sim T^{-\nu}$ as $T \rightarrow 0$.

The same method does not enable one to extract the exponent ν for the $T = 0$ phase transition of the $\pm J$ spin glass[64]. In this case, the probability distribution for the defect energy does not assume the scaling form (3.15) but instead acquires a delta function peak at $\delta E = 0$. This behavior is a peculiar consequence of the discrete nature of the bond randomness, which gives rise to exact energy cancellations and a finite entropy per spin at $T = 0$. The defect energy is also equal to zero in the fully

frustrated model.

A scaling argument by Bray and Moore[100], however, relates the probability distribution for the defect energy to the decay of correlations at $T = 0$. As before, we imagine that on an $L \times L$ lattice, the defect energy measures the effective block coupling J' on length scale L . Let $p(L) \sim L^{-\eta}$ be the fraction of $L \times L$ blocks for which $J' \neq 0$. Assuming that, on length scale L ,

$$\langle \sigma_0 \sigma_L \rangle = \begin{cases} \pm 1 & \text{with probability } p(L) \\ 0 & \text{with probability } 1 - p(L) \end{cases} \quad (3.16)$$

it follows that η is the critical exponent that characterizes the power law decay of correlations $\overline{\langle \sigma_0 \sigma_L \rangle^2}$ at $T = 0$. Plotting $p(L)$ versus L in Figure 3.17, we find $\eta = 0.22 \pm 0.06$ in agreement with several previous results[74, 100, 101].

We also looked at the defect entropy, i.e. the difference in the zero-temperature entropies of systems with periodic and anti-periodic boundary conditions. The mean square defect entropy for the $\pm J$ spin glass is plotted versus lattice size in Figure 3.18. The data is fitted to the scaling law $\sqrt{\delta S_L^2} \sim L^{y_S}$ with $y_S = 0.49 \pm 0.02$. This is curiously close to the result $\delta S_L \sim L^{1/2}$, expected if entropy changes due to reversing the different bonds along the boundary are statistically independent. Figure 3.19 shows the data collapse of the probability distribution to the scaling form

$$P_L(\delta S) = \frac{1}{\sqrt{L}} f_S \left(\frac{\delta S}{\sqrt{L}} \right) \quad (3.17)$$

for lattices of size $L = 8, 12$, and 16 . For purposes of comparison, we also examined

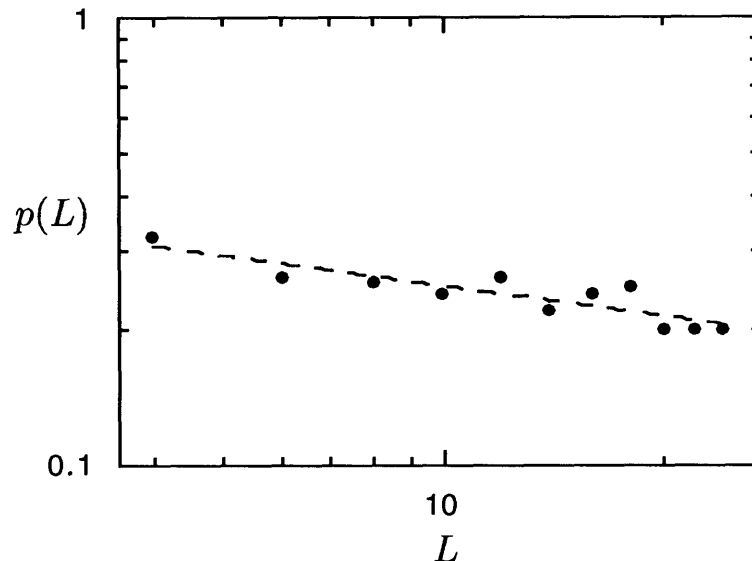


Figure 3.17: Fraction, $p(L)$, of effective block couplings with $J' \neq 0$ versus lattice size. The dashed line shows $p(L) \sim L^{-0.22}$.

the defect entropy of the fully frustrated Ising model. In this case, there is no need to perform a quenched average. Figure 3.20 shows a plot of the defect entropy versus lattice size. The defect entropy in the fully frustrated Ising model approaches a constant value with $1/L^2$ corrections. This behavior is markedly different from the spin glass. We do not know any obvious relation between the finite-size scaling of the defect entropy and other quantities at $T = 0$.

3.4 Extensions

To conclude, we mention a number of possible extensions of this work. An obvious one is treating Ising models with $\pm J$ and/or missing bonds. A large number of random bond and percolation problems fall into this category. Several authors[69, 77, 102]

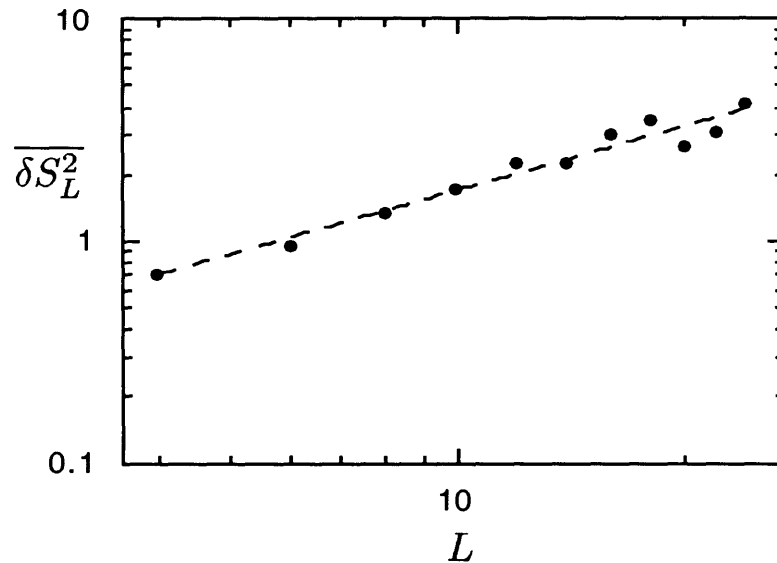


Figure 3.18: Log-log plot of the mean square defect entropy versus lattice size for the $\pm J$ spin glass. The dashed line shows $\overline{\delta S_L^2} \sim L^{0.49}$.

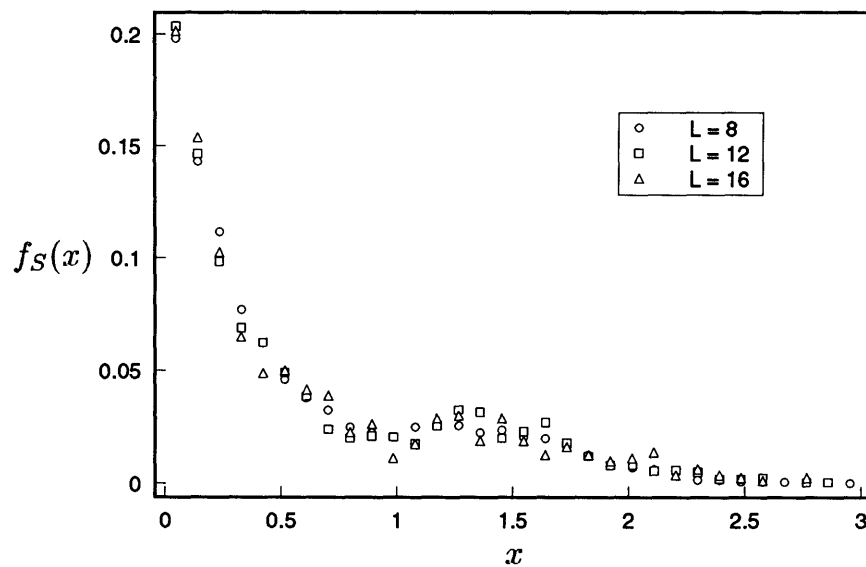


Figure 3.19: Data collapse of the probability distribution for the defect entropy with $x = \delta S/L^{1/2}$.

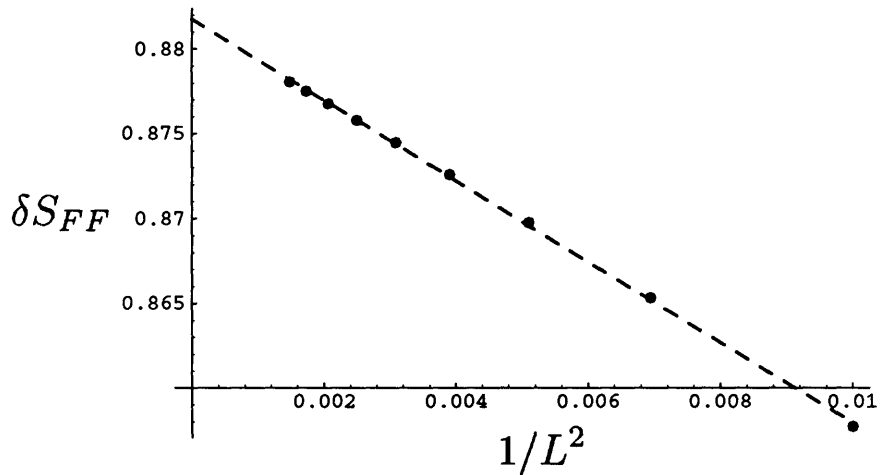


Figure 3.20: Defect entropy versus $1/L^2$ for the fully frustrated model.

have studied the problem that arises when the concentration of frustrated plaquettes x_F is allowed to vary from 0 to 1. The pure Ising model is the case $x_F = 0$; the $\pm J$ spin glass, $x_F = 1/2$; and the fully frustrated model, $x_F = 1$. It is well-known that the ferromagnetism of the ground state disappears for $x_F^c \approx 0.33$. Nevertheless, there remains some question on the nature of the ground state in the vicinity of x_F^c [102]. Besides the transition at x_F^c , there may also be an interesting crossover at $T = 0$ between the spin glass at $x_F = 1/2$ and the fully frustrated model at $x_F = 1$.

If there is a central lesson of our work, however, it is that the Kac-Ward determinant of Eq. (3.10) enables one to compute the partition function of *any* Ising model in polynomial time[103]. As pointed out by Blackman and Poulter[77], the basic method applies to any 2D Ising model with nearest-neighbor interactions, not just to models with $\pm J$ bonds. (Unfortunately, it is not possible to compute the partition function this way in the presence of a magnetic field.) In the case of $\pm J$ bonds, we can perform an integer calculation that yields the exact density of states—a nice bonus. Never-

theless, one should not overlook the potential of computing the determinant of the Kac-Ward matrix for continuously distributed random bonds. In this case, the matrix elements must absorb not only the signs of the bonds but the factors $\tanh(\beta J_{ij})$ as a whole. Given the hopping matrix, however, we can use standard floating-point techniques to evaluate its determinant, and hence the partition function, at any given temperature. Thermodynamic quantities such as the energy and entropy can then be computed by approximating derivatives by numerical differences over small temperature ranges. Determinant formulas also exist for spin-spin correlations in 2D Ising models [22, 104, 105], making it possible to study magnetic susceptibilities. Inoue[78] has performed determinant-based floating-point calculations on 2D Ising models of up to size $L = 100$, a remarkable improvement over what is possible with transfer-matrix techniques. Polynomial-time algorithms based on these determinants should therefore complement well-established methods in the further study of 2D Ising models with quenched randomness.

Acknowledgements

The work in this chapter represents a collaboration with Mehran Kardar. We thank David Huse, Daniel Fisher, Giorgio Parisi, Michael Fisher, Michael Moore, Naoki Kawashima, and Makoto Inoue for useful discussions. This work uses the commercial program MATHEMATICA and the BigNum package developed jointly by INRIA and Digital PRL.

Appendix A. Periodic Boundary Conditions

The algorithm takes as input a set of bonds $\{J_{ij}\}$ on an $L \times L$ square lattice. Given these bonds, the first step of the algorithm is to construct the $4L^2 \times 4L^2$ matrices U_λ that appear in eq. (3.11), with $\lambda \in \{1, 2, 3, 4\}$. To this end, we introduce two transformations that flip the signs of bonds along a single vertical or horizontal strip on the square lattice. Figure 3.21 illustrates the action of these transformations on the set of bonds for a pure Ising ferromagnet, leading to new bond configurations $\{J_{ij}^{(\lambda)}\}$. By definition, $\{J_{ij}^{(1)}\} = \{J_{ij}\}$. The bonds $\{J_{ij}^{(2)}\}$ and $\{J_{ij}^{(3)}\}$ are obtained by flipping a single vertical or horizontal strip of bonds; $\{J_{ij}^{(4)}\}$, by a combination of such flips. The hopping matrices U_λ are given by the elements in Table 3.1, substituting $\{J_{ij}^{(\lambda)}\}$ for $\{J_{ij}\}$. Recall that we have introduced four separate hopping matrices in order to weight correctly those diagrams that loop around the lattice in the horizontal and/or vertical direction. An example of one of these problematic diagrams is shown in Figure 3.2. While no single U_λ matrix assigns the correct weights to these boundary-crossing diagrams, one can verify that the linear combination of determinants in Eq. (3.11) does indeed achieve this goal[94].

Note that the bond-changing transformations in Figure 3.21 do not change the number of frustrated plaquettes on the square lattice. In this sense, we can say that $\{J_{ij}^{(2)}\}$, $\{J_{ij}^{(3)}\}$, and $\{J_{ij}^{(4)}\}$ also represent valid realizations of randomness for the $\pm J$ spin glass on a finite-size lattice. The expressions for the partition functions of these systems are obtained by simply permuting the hopping matrices U_λ . Our algorithm thus computes partition functions for the $\pm J$ spin glass in groups of four. The fact

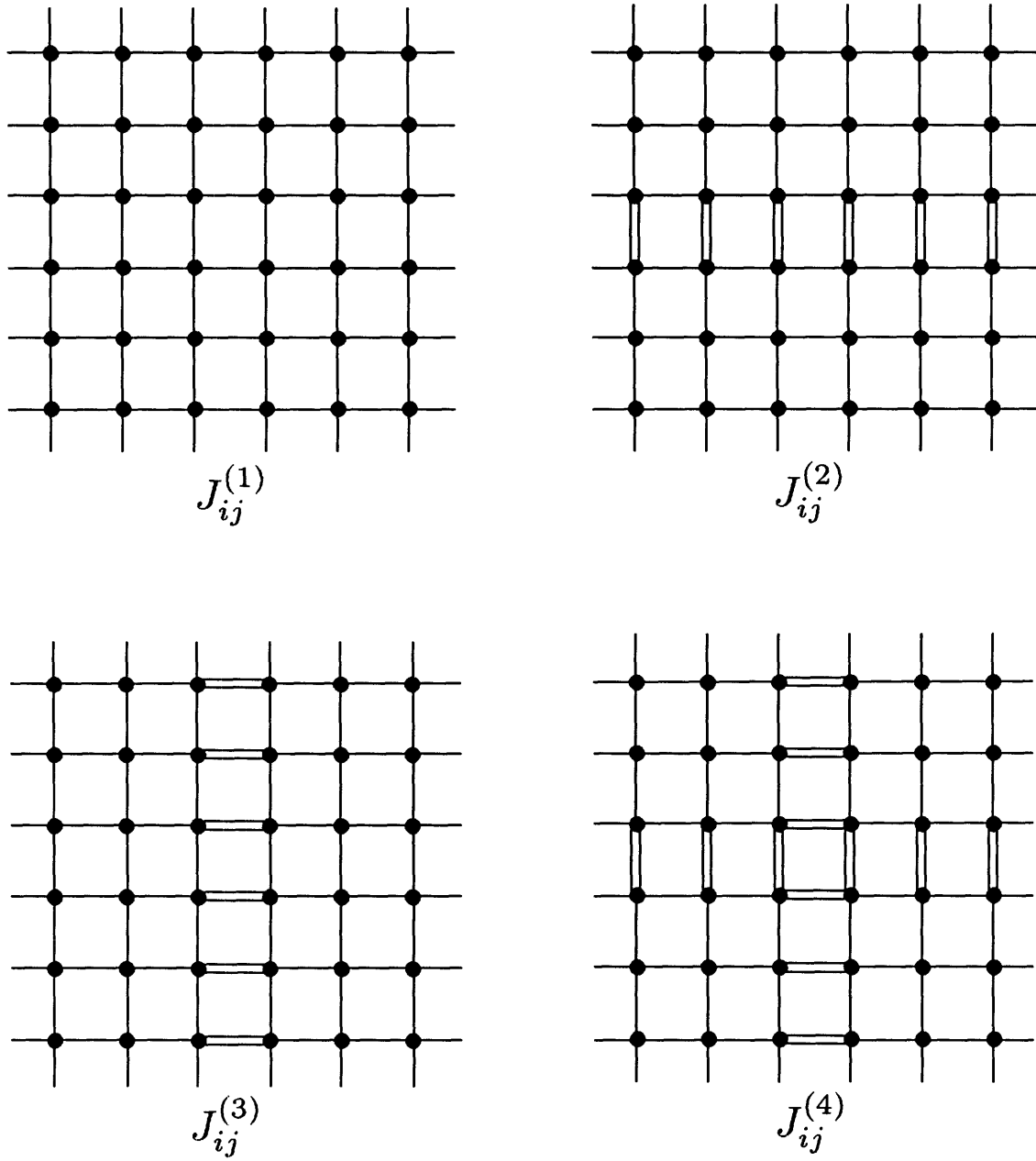


Figure 3.21: Bond transformations for an Ising ferromagnet with $J_{ij} = +J$.

that these four partition functions are related by boundary transformations is useful for studying various properties of the ground state. In particular, the differences in the ground state energies and entropies of these four realizations are precisely the defect energies and entropies defined in Section 3.3.3.

Appendix B. Implementation

This appendix discusses various aspects of the algorithm's implementation, including the handling of large integer arithmetic, sparse matrix multiplication, power series manipulations, and special symmetries. The main loop of the algorithm computes the traces of the matrices $(U_\lambda)^\ell$. An examination of the elements in Table 3.1 reveals an important property of these hopping matrices, namely that one can perform their matrix multiplication using only integer arithmetic. This is done by keeping track of the complex phases $e^{\pm i\pi/4}$ internally and handling their multiplication with lookup tables. Of course, one must also be equipped to deal with the very large integers that appear in the computation of powers of the hopping matrix: the matrix elements of $(U_\lambda)^\ell$ count paths of length ℓ on the square lattice, and the number of these paths grows exponentially in ℓ . We defined special data types to store these integers and perform arithmetic on them; other methods for handling large integers have also appeared in recent computations of series expansions[83]. Storage and time requirements were reduced by multiplying matrices one row at a time and by taking

advantage of the fact that the partial traces satisfy

$$\begin{aligned}\sum_{\mathbf{i}}[(U_\lambda)^\ell]_{i\uparrow,i\uparrow} &= \sum_{\mathbf{i}}[(U_\lambda)^\ell]_{i\downarrow,i\downarrow}, \\ \sum_{\mathbf{i}}[(U_\lambda)^\ell]_{i\leftarrow,i\leftarrow} &= \sum_{\mathbf{i}}[(U_\lambda)^\ell]_{i\rightarrow,i\rightarrow}.\end{aligned}$$

In addition, we exploited the sparseness of the hopping matrices to multiply them in $O(N^2)$ rather than $O(N^3)$ steps. Computing the required powers of the hopping matrices U_λ thus takes a total of $O(N^3)$ steps. Note that the computation of these traces can in principle be done by $8N$ parallel processors, one for each bond on the λ th square lattice; obviously, this would yield a substantial improvement in performance.

Equipped with the traces of the matrices $(U_\lambda)^\ell$, we proceed to compute the component partition functions Z_λ , as defined in Eq. (3.11), using a sequence of power series manipulations. The component partition functions obey

$$\ln Z_\lambda = N \ln 2 + N \ln[\cosh(\beta J)] - \sum_{\ell=0}^{\infty} \frac{1}{\ell} \tanh^\ell(\beta J) \text{tr}[(U_\lambda)^\ell]. \quad (3.18)$$

We obtain the coefficients of the high temperature series for Z_λ by exponentiating the right hand side of Eq. (3.18) and grouping terms in powers of $\tanh(\beta J)$. Combining the component partition functions as $Z = -Z_1 + Z_2 + Z_3 + Z_4$ leads to an expression for the partition function in the form of Eq. (3.8),

$$Z = 2^N \cosh^{2N}(\beta J) \sum_{\ell=0}^{2N} A_\ell \tanh^\ell(\beta J).$$

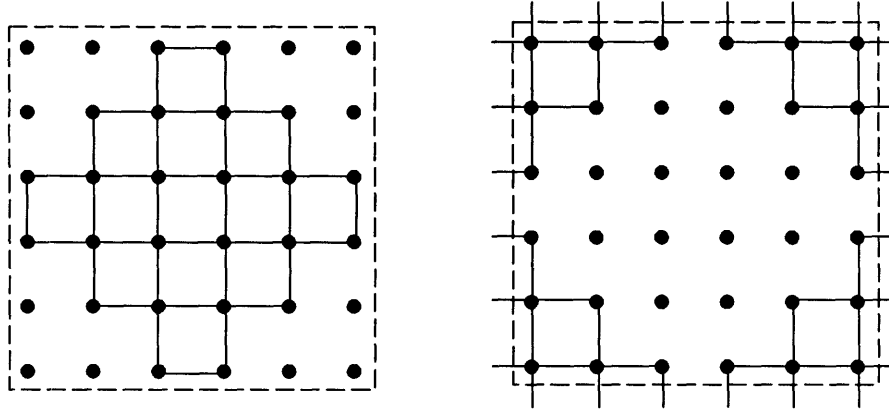


Figure 3.22: Complementary graphs on the square lattice with periodic boundary conditions. This complementarity gives rise to a symmetry $A_\ell = A_{2N-\ell}$ in the coefficients of the high-temperature expansion.

On a square lattice with periodic boundary conditions, the high temperature series for Z is a polynomial of order $2N$ in $\tanh(\beta J)$. With $\pm J$ bonds, however, the coefficient of $\tanh^\ell(\beta J)$ is equal, up to a possible sign, to that of $\tanh^{(2N-\ell)}(\beta J)$ due to a complementarity in the diagrammatic expansion (see Figure 3.22). The coefficients have opposite sign if the product of all bonds on the square lattice, $\prod J_{ij}$, is negative; otherwise, they are equal. Due to this symmetry, only the traces of $(U_\lambda)^\ell$ with $\ell \leq N$ are necessary to compute the complete high temperature series for Z .

The final task is to extract the density of states from the integer coefficients A_ℓ in eq. (3.8). This is done by simply expanding $\cosh(\beta J)$ and $\tanh(\beta J)$ in terms of $e^{-\beta J}$. The amount of algebra can be reduced by first cancelling powers of $\cosh^2(\beta J)$ and $[1 - \tanh^2(\beta J)]$ that factorize the high temperatures series. The final result,

$$Z = \sum_E g(E) e^{-\beta E}, \quad (3.19)$$

yields the density of states $g(E)$ as a function of energy.

Chapter 4

Learning in Boltzmann Trees

4.1 Introduction and Summary

Statistical mechanics is a powerful tool for analyzing systems with many degrees of freedom. In the last decade, a number of workers have used statistical mechanics to study problems outside the traditional domain of physics. One area of intense research has been the field of parallel distributed processing[27], or neural computation[8]. A fundamental goal in this field is to understand the computational abilities of networks of simple processing units. Of particular interest to physicists are networks whose units obey stochastic update rules, similar to Ising spins in contact with a heat bath. One can use statistical mechanics to analyze the equilibrium behavior of these networks and to design learning algorithms that optimize their performance.

The behavior of a network depends on the values of its connections, or weights. Programming the network to perform a certain task is done by assigning appropriate values to these weights. For some tasks, such as the storage of associative

memories[12], it is possible to choose the correct weights a priori; for most tasks, however, this is not the case. Instead, one must conduct a systematic search through weight space to find an appropriate set of values. During this training phase, the weights are iteratively adjusted in response to evaluations of the network's performance. Training is stopped when the network has learned to reproduce the desired behavior.

To illustrate this, let us return to the simple binary perceptron, introduced in Chapter I. This network has input units $\{S_i\}$, weights $\{J_i\}$, offset h , and output S . The input-output mapping is $S = \text{sgn}(\sum J_i S_i + h)$, as shown in Figure 1.1. Without loss of generality, we take the bias h to be zero, since its effect can be incorporated by adding an extra unit to each input pattern. In the problem of learning from examples, we are given a set of input patterns $\{S_i(\mu)\}_{\mu=1}^p$ and target outputs $\{S^*(\mu)\}_{\mu=1}^p$; the goal is to find a set of weights $\{J_i\}$ that correctly classifies all the input patterns. This problem is an example of *supervised* learning—supervised, in the sense that an imaginary teacher has explicitly provided the perceptron with p target outputs for the patterns $\{S_i(\mu)\}$.

As discussed in Chapter I, a solution exists if and only if the patterns in the training set are linearly separable. Let us assume this to be the case. For binary outputs ± 1 , the number of mistakes on the training set is equal to

$$\epsilon = \frac{1}{4} \sum_{\mu=1}^p \left\{ S^*(\mu) - \text{sgn} \left[\sum_i J_i S_i(\mu) \right] \right\}^2 ; \quad (4.1)$$

eq. (4.1) is called the cost function for the perceptron. Rosenblatt[37] devised a

learning algorithm for the perceptron that iteratively adjusts the weights $\{J_i\}$ to decrease the number of mistakes on the training set. The basic idea is to cycle through the examples in the training set, adapting the weights whenever the perceptron makes an error. Thus, for each input pattern, the weights of the perceptron are changed by an amount

$$\Delta J_i = \eta \left\{ S^*(\mu) - \text{sgn} \left[\sum_i J_i S_i(\mu) \right] \right\} S_i(\mu),$$

where $\eta > 0$ is the learning rate. This process continues until the perceptron correctly classifies all the input patterns in the training set. Provided that these examples are linearly separable, the perceptron learning algorithm will converge to a solution in a finite number of steps[37, 38].

The restriction to linearly separable problems limits the usefulness of simple perceptrons for real-world tasks. As pointed out in the Chapter I, multilayer networks with hidden units do not share this fundamental limitation. The first algorithms for training multilayer networks were reported in the early 1980s; the result was an explosion of activity in the field of neural networks. The general idea behind many of these algorithms is to minimize a cost function that measures the discrepancy between the observed and desired performance of the network. In this chapter, we will explore a particular type of network based on the statistical mechanics of Ising spin systems. These networks, known as Boltzmann machines[39], have several compelling virtues. Unlike simple perceptrons, they can solve problems that are not linearly separable. The learning rule, simple and locally based, lends itself to massive parallelism. The theory of Boltzmann learning, moreover, has a solid foundation in statistical mechan-

ics.

Unfortunately, Boltzmann machines—as originally conceived—also have some serious drawbacks. In practice, they are extremely slow. The weight adjustments are based on stochastic correlations that are difficult to compute. The general procedure for estimating them is Monte Carlo simulation[109], combined with the method of simulated annealing[11]. This approach, though effective, entails a great deal of computation. Finally, compared to back-propagation networks[110], where weight updates are computed by the chain rule, Boltzmann machines lack a certain degree of exactitude. Monte Carlo estimates of stochastic averages are not sufficiently accurate to permit further refinements to the learning rule, such as quasi-Newton or conjugate-gradient techniques[111].

There have been efforts to overcome these difficulties. Peterson and Anderson[112] introduced a mean-field version of the original Boltzmann learning rule. For many problems, this approximation works surprisingly well[113], so that mean-field Boltzmann machines learn much more quickly than their stochastic counterparts. Under certain circumstances, however, the approximation breaks down, and the mean-field learning rule works badly if at all[114]. Another approach[115] is to focus on Boltzmann machines with architectures simple enough to permit exact computations. Learning then proceeds by straightforward gradient descent on the Boltzmann machine cost function[116], without the need for simulated or mean-field annealing. Hopfield[115] wrote down the complete set of learning equations for a Boltzmann machine with one layer of non-interconnected hidden units. Freund and Haussler[117] derived the analogous equations for the problem of unsupervised learning.

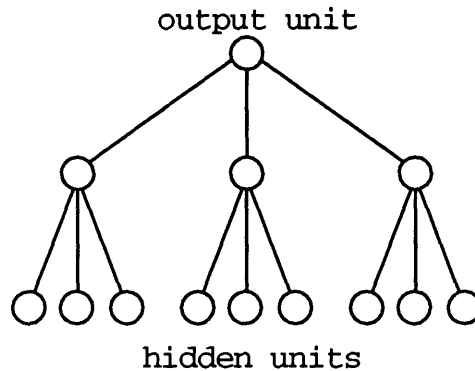


Figure 4.1: Boltzmann tree with two layers of hidden units. The input units (not shown) are fully connected to all the units in the tree.

In this chapter, we pursue this strategy further, concentrating on the case of supervised learning. In section 4.2, we review the Boltzmann learning algorithm and its implementation for stochastic and mean-field Boltzmann machines. In section 4.3, we exhibit a large family of architectures for which it is possible to implement the Boltzmann learning rule in an exact way. The networks in this family have a hierarchical structure with tree-like connectivity. In general, they can have one or more layers of hidden units¹. We call them Boltzmann trees; an example is shown in Figure 4.1. We use a decimation technique from statistical physics to compute the averages in the Boltzmann learning rule. After describing the method, we give results on the problems of N -bit parity and the detection of hidden symmetries[118]. We also compare the performance of deterministic and true Boltzmann learning. Finally, in section 4.5, we discuss a number of possible extensions to our work.

¹The hidden and output units form a tree; there are no connections between the input units since they are not allowed to equilibrate. See Section 4.2.

4.2 Boltzmann Machines

We briefly review the learning algorithm for the Boltzmann machine[8]. The Boltzmann machine is a recurrent network with binary units $S_i = \pm 1$ and symmetric weights $w_{ij} = w_{ji}$. Each configuration of units in the network represents a state of energy

$$\mathcal{H} = - \sum_{ij} w_{ij} S_i S_j. \quad (4.2)$$

The network operates in a stochastic environment in which states of lower energy are favored. The units in the network change states with probability

$$P(S_i \rightarrow -S_i) = \frac{1}{1 + e^{\Delta\mathcal{H}/T}}. \quad (4.3)$$

Once the network has equilibrated, the probability of finding it in a particular state obeys the Boltzmann distribution from statistical mechanics:

$$P_B(S_1, \dots, S_N) = \frac{1}{Z} e^{-\mathcal{H}/T}. \quad (4.4)$$

The partition function $Z = \sum e^{-\mathcal{H}/T}$ is the weighted sum over states needed to normalize the Boltzmann distribution. The temperature T determines the amount of noise in the network; as the temperature is decreased, the network is restricted to states of lower energy.

We consider a network with input units (I), hidden units (H), and output units (O). The problem to be solved is one of supervised learning. Input patterns are selected

from a training set with probability $P^*(I_\mu)$. Likewise, target outputs are drawn from a probability distribution $P^*(O_\nu|I_\mu)$; the notation $P^*(O_\nu|I_\mu)$ denotes the conditional probability that the desired output is O_ν given that the input pattern is I_μ . The goal is to teach the network the desired associations. Both the input and output patterns are binary. The μ th example in the training set is said to be learned if after clamping the input units to pattern I_μ and waiting for the network to equilibrate, the output units are in the states O_ν with probability $P^*(O_\nu|I_\mu)$. The input-output mapping performed by the network clearly depends on the network connectivity, as well as the values of the weights².

A suitable cost function for this supervised learning problem is

$$E = \sum_{\mu} P^*(I_{\mu}) \sum_{\nu} P^*(O_{\nu}|I_{\mu}) \ln \left[\frac{P^*(O_{\nu}|I_{\mu})}{P(O_{\nu}|I_{\mu})} \right] \quad (4.5)$$

where $P^*(O_\nu|I_\mu)$ and $P(O_\nu|I_\mu)$ are the desired and observed probabilities that the output units have pattern O_ν when the input units are clamped to pattern I_μ . This cost function, known as the relative entropy, measures the difference between the desired and observed probability distributions. The relative entropy between two probability distributions is zero if the distributions are the same; otherwise, it is strictly positive. An application of the relative entropy cost function, familiar to physicists, is the mean-field approximation discussed at the end of this section.

The Boltzmann learning algorithm attempts to minimize the cost function (4.5)

²Note that during the operation of the network, only the hidden and output units are allowed to equilibrate, while the input units remain clamped. Weights between input units therefore have no effect on the operation of the network. Without loss of generality, we take these weights to be zero.

by gradient descent. The calculation of the gradients in weight space is made possible by the general relation

$$\frac{\partial(\ln Z)}{\partial w_{ij}} = \frac{1}{T} \langle S_i S_j \rangle, \quad (4.6)$$

where brackets $\langle \dots \rangle$ indicate expectation values over the Boltzmann distribution. The cost function (4.5) depends on the weights w_{ij} through the conditional probability

$$P(O_\nu | I_\mu) = \frac{Z_{I_\mu, O_\nu}}{Z_{I_\mu}}, \quad (4.7)$$

where the subscripts on Z indicate that the input and/or output units are clamped and not allowed to equilibrate. Thus,

$$Z_{I_\mu, O_\nu} = \sum_{\{H\}} e^{-\beta \mathcal{H}} \quad (4.8)$$

only traces over the states of the hidden units (H), while

$$Z_{I_\mu} = \sum_{\{O, H\}} e^{-\beta \mathcal{H}} \quad (4.9)$$

traces over the states of the hidden (H) and output (O) units. The Boltzmann learning rule minimizes the relative entropy cost function by gradient descent:

$$\begin{aligned} \Delta w_{ij} &= -\eta \frac{\partial E}{\partial w_{ij}} \\ &= \frac{\eta}{T} \sum_{\mu} P^*(I_\mu) \sum_{\nu} P^*(O_\nu | I_\mu) [\langle S_i S_j \rangle_{I, O}^{\mu, \nu} - \langle S_i S_j \rangle_I^{\mu}]. \end{aligned} \quad (4.10)$$

The gradients in weight space depend on two sets of correlations—one in which the output units are clamped to their desired targets O_ν , the other in which they are allowed to equilibrate. In both cases, the input units are clamped to the pattern I_μ being learned. The differences in these correlations³, averaged over the examples in the training set, yield the weight updates Δw_{ij} . The parameter η sets the learning rate. For very large training sets with redundant input patterns, it may be desirable to update the weights on a more frequent basis. An “on-line” version of Boltzmann learning is obtained by foregoing the average over input patterns and updating the weights after each example.

The main drawback of Boltzmann learning is that, in most networks, it is not possible to compute the gradients in weight space directly. Instead, one must resort to estimating the correlations $\langle S_i S_j \rangle$ by Monte Carlo simulation[109]. In this procedure, the units are stochastically updated using eq. (4.3), and time-averaged correlations are measured over many sweeps through the network. The time to reach equilibrium is greatly reduced by starting the simulation at high temperatures and gradually lowering the temperature to the desired value. This method of simulated annealing[11] leads to acceptable estimates of stochastic averages but has the disadvantage of being very computation-intensive.

A deterministic version of the algorithm[112] was proposed to speed up learning. It makes the approximation $\langle S_i S_j \rangle \approx \langle S_i \rangle \langle S_j \rangle$ in the learning rule and estimates the magnetizations $\langle S_i \rangle$ within the mean-field approximation. There are several ways to

³Weights from input units are updated by a simplified Boltzmann learning rule; if A is an input unit, and B a hidden or output unit, then $\langle AB \rangle_I^\mu = \pm \langle B \rangle_I^\mu$, where the sign is determined by the clamped value of A in the μ th pattern in the training set.

derive the mean-field equations; we choose one that illustrates the relative entropy cost function. The key idea behind the mean-field approximation is to replace the Boltzmann distribution (4.4) by a simpler one that factors into N independent terms:

$$P_{MF}(S_1, \dots, S_N) = \prod_{i=1}^N \left(\frac{1 + m_i S_i}{2} \right). \quad (4.11)$$

Within this approximation, one has $\langle S_i \rangle = m_i$ and $\langle S_i S_j \rangle = m_i m_j + \delta_{ij}(1 - m_i^2)$. To make the approximation a useful one, the values of m_i are chosen so that P_{MF} is as close as possible to the Boltzmann distribution P_B . This is done by minimizing the relative entropy between the two distributions,

$$\begin{aligned} \Psi(m_1, \dots, m_N) &= \sum_{\{S_i = \pm 1\}} P_{MF}(S_1, \dots, S_N) \ln \left[\frac{P_{MF}(S_1, \dots, S_N)}{P_B(S_1, \dots, S_N)} \right] \\ &= \sum_i \left[\left(\frac{1 + m_i}{2} \right) \ln \left(\frac{1 + m_i}{2} \right) + \left(\frac{1 - m_i}{2} \right) \ln \left(\frac{1 - m_i}{2} \right) \right] \\ &\quad + \frac{1}{T} \sum_{ij} w_{ij} m_i m_j + \ln Z. \end{aligned} \quad (4.12)$$

The minimum relative entropy is found by setting the derivatives with respect to m_i to zero. A straightforward calculation shows that the optimal values of m_i satisfy

$$m_i = \tanh \left(\frac{1}{T} \sum_j w_{ij} m_j \right). \quad (4.13)$$

These are the standard mean-field equations, which can also be derived by replacing the units on the right hand side of eq. (4.3) by their mean values.

Mean-field Boltzmann machines solve these equations by iteration, combined when

necessary with an annealing process (i.e. a gradual lowering of T). The resulting values of m_i are used to estimate the correlations $\langle S_i S_j \rangle$ and update the weights according to the Boltzmann learning rule. On simple benchmark problems, Peterson and Anderson[112] found mean-field Boltzmann machines to be 10-30 times faster than their stochastic counterparts. The problem remains, however, that the mean-field approximation loses its validity when the units in the network are strongly correlated with one another. A recent study by Galland[114] suggests that mean-field learning breaks down in networks with more than one layer of hidden units.

4.3 Boltzmann Trees

Clearly, the ideal algorithm would be one that computes expectation values exactly and does not involve the added complication of annealing. In this section, we investigate a large family of networks amenable to exact computations of this sort. These networks, which we call Boltzmann trees, are simply Boltzmann machines whose hidden and output units have a special hierarchical organization. There are no restrictions on the input units, and in general, we will assume them to be fully connected to the rest of the network. For convenience, we will focus on the case of one output unit; an example of such a Boltzmann tree is shown in Figure 4.1. Modifications to this basic architecture and the generalization to many output units will be discussed later.

The key technique to compute partition functions and expectation values in these trees is known as decimation[19, 119]. The idea behind decimation is the following.

Consider three units connected in series, as shown in Figure 4.2a. Though not directly connected, the end units S_1 and S_2 have an effective interaction that is mediated by the middle one S . Define the temperature-rescaled weights $J_{ij} \equiv w_{ij}/T$. We claim that the combination of the two weights J_1 and J_2 in series has the same effect as a single weight J . Replacing the weights in this way, we have integrated out, or “decimated”, the degree of freedom represented by the intermediate unit. To derive an expression for J , we require that the units S_1 and S_2 in both systems obey the same Boltzmann distribution. This will be true if

$$\sum_{S=\pm 1} e^{J_1 S_1 S + J_2 S S_2} = \sqrt{C} e^{J S_1 S_2}, \quad (4.14)$$

where C is a constant prefactor, independent of S_1 and S_2 . Enforcing equality for the possible values of $S_1 = \pm 1$ and $S_2 = \pm 1$, we obtain the constraints

$$\sqrt{C} e^{\pm J} = 2 \cosh(J_1 \pm J_2).$$

It is straightforward to eliminate C and solve for the effective weight J . Omitting the algebra, we find

$$\tanh J = \tanh J_1 \tanh J_2. \quad (4.15)$$

Choosing J in this way, we ensure that all expectation values involving S_1 and/or S_2 will be the same in both systems.

Decimation is a technique for combining weights “in series”. The much simpler case of combining weights “in parallel” is illustrated in Figure 4.2b. In this case, the

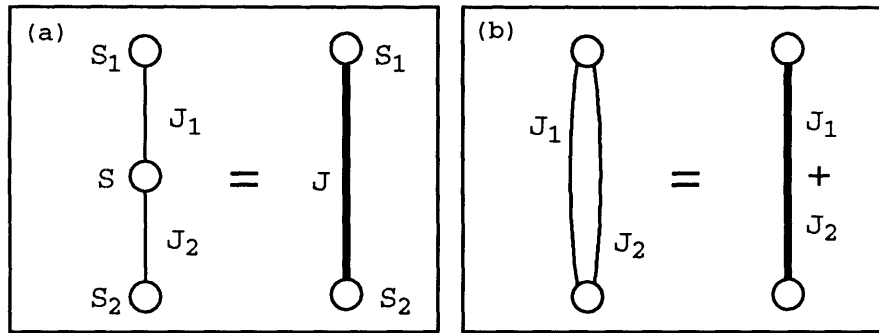


Figure 4.2: (a) Combining weights in series: the effective interaction between units S_1 and S_2 is the same as if they were directly connected by weight J , where $\tanh J = \tanh J_1 \tanh J_2$. (b) Combining weights in parallel: the effective weight is simply the additive sum. The same rules hold if either of the end units is clamped.

effective weight is simply the additive sum of J_1 and J_2 , as can be seen by appealing to the energy function of the network, eq. (4.2). Note that the rules for combining weights in series and in parallel remain valid if either of the end units S_1 or S_2 happen to be clamped. They also hold locally for weight combinations that are embedded in larger networks. The rules have simple analogs in other types of networks (e.g. the law for combining resistors in electric circuits). Indeed, the strategy for exploiting these rules is a familiar one. Starting with a complicated network, we iterate the rules for combining weights until we have a simple network whose properties are easily computed. Clearly, the rules do not make all networks tractable; networks with full connectivity between hidden units, for example, cannot be systematically reduced. Hierarchical networks with tree-like connectivity, however, lend themselves naturally to these types of operations.

Let us see how we can use these rules to implement the Boltzmann learning rule in an exact way. Consider the two-layer Boltzmann tree in Figure 4.1. The effect of clamping the input units to a selected pattern is to add a bias to each of the

units in the tree, as in Figure 4.3a. Note that these biases depend not only on the input weights, but also on the pattern distributed over the input units. Having clamped the input units, we must now compute expectation values. For concreteness, we consider the case where the output unit is allowed to equilibrate. Correlations between adjacent units are computed by decimating over the other units in the tree; the procedure is illustrated in Figure 4.3b for the lower leftmost hidden units. The final, reduced network consists of the two adjacent units with weight J and effective biases (h_1, h_2) . A short calculation gives

$$\langle S_1 S_2 \rangle = \frac{e^J \cosh(h_1 + h_2) - e^{-J} \cosh(h_1 - h_2)}{e^J \cosh(h_1 + h_2) + e^{-J} \cosh(h_1 - h_2)}. \quad (4.16)$$

The magnetization of a tree unit can be computed in much the same way. We combine weights in series and parallel until only the unit of interest remains, as in Figure 4.3c. In terms of the effective bias h , we then have the standard result

$$\langle S_1 \rangle = \tanh h. \quad (4.17)$$

The rules for combining weights thus enable us to compute expectation values without enumerating the $2^{13} = 8192$ possible configurations of units in the tree. To compute the correlation $\langle S_1 S_2 \rangle$ for two adjacent units in the tree, one successively removes all “outside” fluctuating units until only units S_1 and S_2 remain. To compute the magnetization $\langle S_1 \rangle$, one removes unit S_2 as well.

Implementing these operations on a computer is relatively straightforward, due

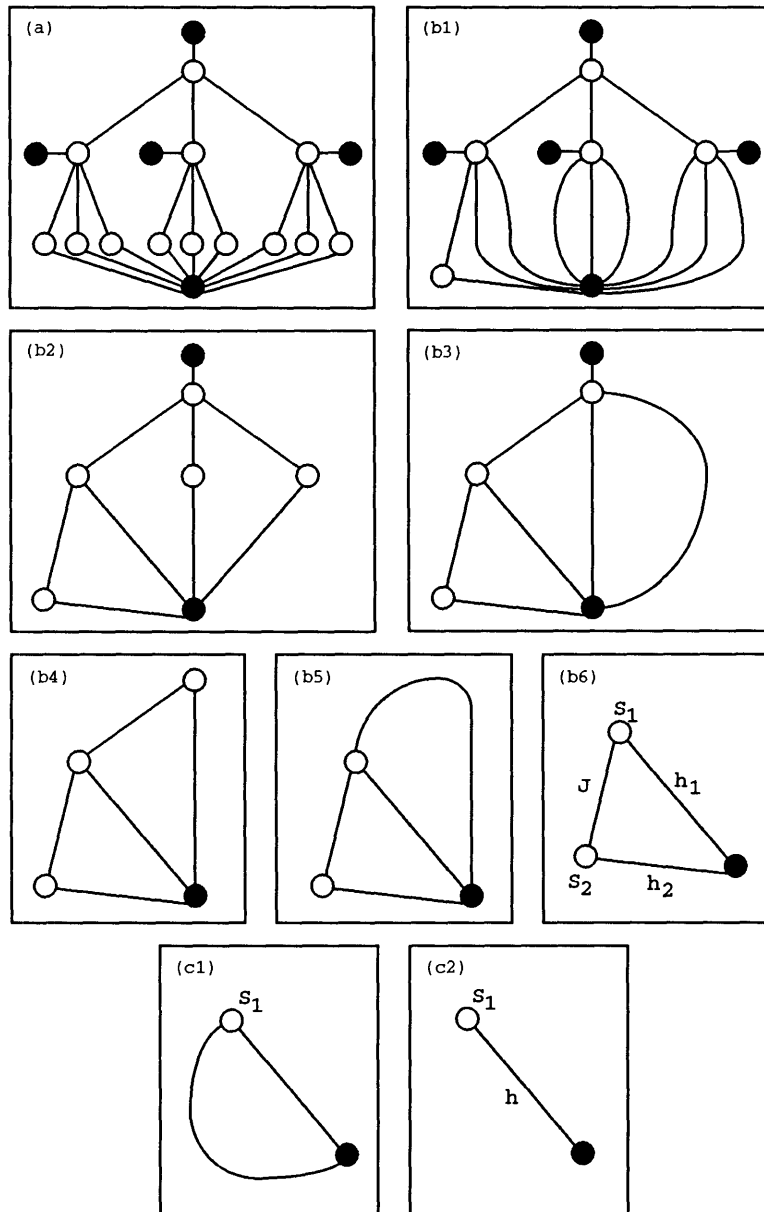


Figure 4.3: Reducing Boltzmann trees by combining weights in series and parallel. Solid circles represent clamped units. (a) Effect of clamping the input units to a selected pattern. (b) Computing the correlation between adjacent units. (c) Computing the magnetization of a single unit.

to the hierarchical organization of the output and hidden units. The entire set of correlations and magnetizations can be computed by making two recursive sweeps through the tree, storing effective weights as necessary to maximize the efficiency of the algorithm. Having to clamp the output unit to the desired target does not introduce any difficulties. In this case, the output unit merely contributes (along with the input units) to the bias on its derivative units. Again, we use recursive decimation to compute the relevant stochastic averages. We are thus able to implement the Boltzmann learning rule in an exact way.

4.4 Results

We tested Boltzmann trees on two familiar problems: N -bit parity and the detection of hidden symmetries[118]. We hope our results demonstrate not only the feasibility of the algorithm, but also the potential of exact Boltzmann learning. Table 4.1 shows our results on the N -bit parity problem, using Boltzmann trees with one layer of hidden units. In each case, we ran the algorithm 1000 times. All 2^N possible input patterns were included in the training set. A success indicates that the tree learned the parity function in less than e_{max} epochs, where an epoch is defined as one pass through the training set. We also report the average number of epochs e_{avg} per successful trial; in these cases, training was stopped when $P(O^*|I_\mu) \geq 0.9$ for each of the 2^N inputs, with $O^* = \text{parity}(I_\mu)$. The results show Boltzmann trees to be competitive with standard back-propagation networks[120].

We also tested Boltzmann trees on the problem of detecting hidden symmetries.

N	hidden units	e_{max}	success %	e_{avg}
2	1	50	97.2 (89.3)	25.8
3	1	250	96.1 (88.5)	42.1
4	3	1000	95.1 (69.2)	281.1
5	4	1000	92.9 (84.2)	150.0

Table 4.1: Boltzmann tree performance on N -bit parity. The results in parentheses are for mean-field learning.

In the simplest version of this problem, the input patterns are square pixel arrays which have mirror symmetry about a fixed horizontal or vertical axis (but not both). Figure 4.4 shows examples of both types of patterns. We used a two-layer tree with the architecture shown in Figure 4.1 to detect these symmetries in 10×10 square arrays. The network learned to differentiate the two types of patterns from a training set of 2000 labelled examples. After each epoch, we tested the network on a set of 200 unknown examples. The performance on these patterns measures the network’s ability to generalize to unfamiliar inputs. The results, averaged over 100 separate trials, are shown in Figure 4.5. After 100 epochs, average performance was over 95% on the training set and over 85% on the test set.

Finally, we investigated the use of the deterministic[112], or mean-field, learning rule in Boltzmann trees. We repeated our experiments, substituting $\langle S_i \rangle \langle S_j \rangle$ for $\langle S_i S_j \rangle$ in the update rule. Note that we computed the magnetizations $\langle S_i \rangle$ exactly using decimation. In fact, in most deterministic Boltzmann machines, one does not compute the magnetizations exactly, but estimates them within the mean-field approximation. Such networks therefore make two approximations—first, that $\langle S_i S_j \rangle \approx \langle S_i \rangle \langle S_j \rangle$ and second, that $\langle S_i \rangle \approx \tanh(\sum J_{ij} \langle S_j \rangle + h_i)$. Our results speak to

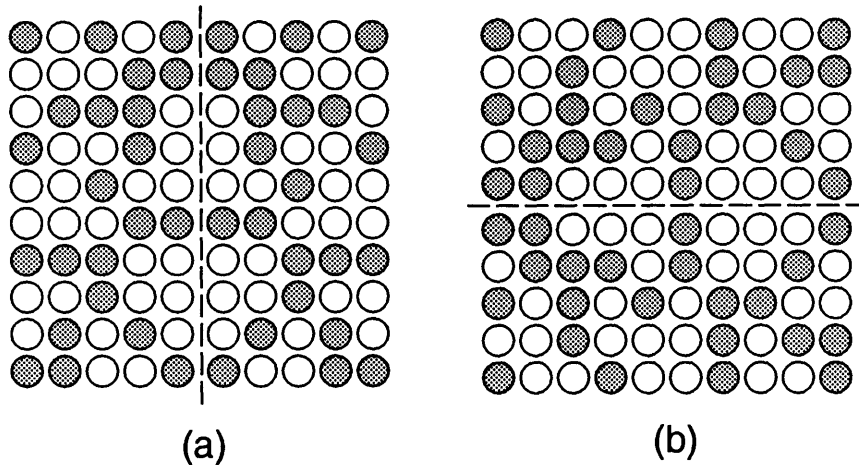


Figure 4.4: Input patterns for the problem of detecting hidden symmetries: (a) vertical mirror symmetry (b) horizontal mirror symmetry.

the first of these approximations. At this level alone, we find that exact Boltzmann learning is perceptibly faster than mean-field learning. On one problem in particular, that of $N = 4$ parity (see Table I), the difference between the two learning schemes was quite pronounced.

4.5 Extensions

In conclusion, we mention several possible extensions to the work in this chapter. Clearly, a number of techniques used to train back-propagation networks, such as conjugate-gradient and quasi-Newton methods[111], could also be used to accelerate learning in Boltzmann trees. In this chapter, we have considered the basic architecture in which a single output unit sits atop a tree of one or more hidden layers. Depending on the problem, a variation on this architecture may be more appropriate. The network must have a hierarchical organization to remain tractable; within this framework, however, the algorithm permits countless arrangements of hidden and

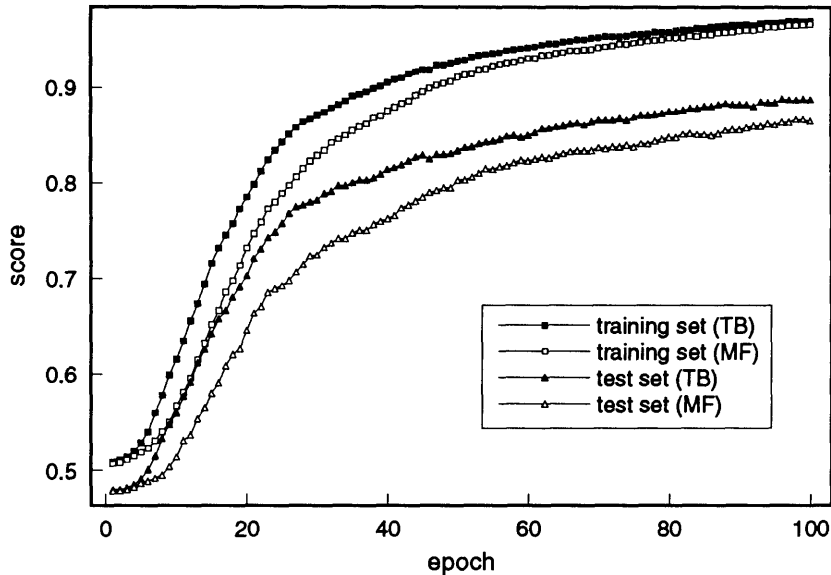


Figure 4.5: Results on the problem of detecting hidden symmetries for true Boltzmann (TB) and mean-field (MF) learning.

output units. In particular, a tree can have one or more output units, and these output units can be distributed in an arbitrary way throughout the tree. One can incorporate certain intralayer connections into the tree at the expense of introducing a slightly more complicated decimation rule, valid when the unit to be decimated is biased by a connection to an additional clamped unit. There are also decimation rules for q -state (Potts) units[19], with $q > 2$.

The algorithm for Boltzmann trees raises a number of interesting questions. Some of these involve familiar issues in neural network design—for instance, how to choose the number of hidden layers and units. We would also like to characterize the types of learning problems best-suited to Boltzmann trees. A recent study by Galland[114] suggests that mean-field learning has trouble in networks with several layers of hidden units and/or large numbers of output units. Boltzmann trees with exact Boltzmann learning may present a viable option for problems in which the basic as-

sumption behind mean-field learning—that the units in the network can be treated independently—does not hold. We know of constructive algorithms[121] for feed-forward nets that yield tree-like solutions; an analogous construction for Boltzmann machines has obvious appeal, in view of the potential for exact computations. Finally, the tractability of Boltzmann trees is reminiscent of the tractability of tree-like belief networks, proposed by Pearl[122, 123]; more sophisticated rules for computing probabilities in belief networks[124] may have useful counterparts in Boltzmann machines. These issues and others are left for further study.

Acknowledgements

The work in this chapter represents a collaboration with Michael Jordan. We thank Mehran Kardar for useful discussions.

Chapter 5

Conclusion

This thesis has investigated the potential for exact computations in the statistical mechanics of disordered systems. Three types of disordered systems were examined: directed waves in random media, the 2d $\pm J$ Ising spin glass, and tree-like Boltzmann machines. In conclusion, we mention some recent works motivated by the ideas in this thesis, as well as some open problems for future research.

In Chapter II, we introduced an S -matrix model for the unitary propagation of directed waves in random media. The model, based on a discretization of the Feynman path integral, reduced the calculation of disorder-averaged quantities to a sum over directed paths on the square lattice. We computed asymptotic scaling laws for the diffusion of the beam width and the wandering of the beam center in two or more dimensions. These scaling laws had a simple interpretation in terms of the properties of random walks.

Recently, a number of authors have obtained additional results from the random S -matrix model. Following our work, Friedberg and Yu[126] calculated the leading terms

in the scaling laws for the beam center in $d \geq 2$ and also the next-order corrections. The analytical results were in agreement with those presented in Chapter II. Cule and Shapir[125] extended the methods of this thesis to compute the higher moments of the probability distribution for directed waves in random media. If this probability distribution is multifractal, as claimed by Bouchaud et al[59], the higher moments should obey new scaling laws whose exponents are not simply related to those of the lower moments. Within the framework of the S -matrix model, Cule and Shapir did not find evidence for multifractal scaling, suggesting that certain forms of scaling behavior may be sensitive to details of the unitary time evolution.

Future investigations with the S -matrix model should lead to a greater understanding of random Schrödinger equations. The current situation is reminiscent of the outlook several years ago for directed polymers[44]. Transfer-matrix studies of the polymer problem clarified many important issues, lending numerical support to exact solutions in $d = 2$ and providing insights into $d > 2$ where exact solutions do not exist. S -matrix models of random Schrödinger equations should be similarly useful for studying wave propagation and localization in disordered media. The prospect of performing exact averages over disorder makes S -matrix models particularly attractive.

Sums over paths appeared in a rather different context in Chapter III. Here, we introduced an exact integer algorithm to compute the density of states for the $2d \pm J$ Ising spin glass. The algorithm was based on the Kac-Ward method[23] for summing the high-temperature series of the partition function. The diagrammatic representation of this series was used to rewrite the calculation of the partition function as

a sum over closed loops on the square lattice. Thus, like the S -matrix model for directed waves, the algorithm for the spin glass was motivated by reformulating the problem as a sum over paths.

Our method for computing the partition function of the 2d $\pm J$ spin glass had three compelling features. First, the computation time was polynomial in the lattice size. Second, the results were not susceptible to floating-point error. Third, the main routine of the algorithm permitted a trivial parallelization. This last feature of the algorithm has yet to be exploited, but it may be the key to resolving several open questions. Data on larger system sizes is necessary, for example, to determine whether the correlation length diverges algebraically or exponentially in $d = 2$. The nature of this divergence may provide clues to the lower critical dimension of the $\pm J$ spin glass.

The prospect of exact polynomial-time algorithms raises some interesting historical parallels. Fifteen years ago, it was widely believed, on the basis of Monte Carlo simulations, that there was a finite-temperature phase transition in $d = 2$ spin glasses. This view was refuted by Morgenstern and Binder[69], who showed using transfer-matrix results that these simulations did not report true equilibrium behavior at low temperatures. Since its introduction, the numerical transfer-matrix technique has become a widespread tool for investigating spin glasses. Our algorithm makes clear once again the value of exact computations in the study of spin glasses. Polynomial-time algorithms should complement transfer-matrix studies and Monte Carlo simulations in future work. As noted in Chapter III, an efficient parallel implementation of our method should permit the study of larger systems than can be reliably tackled by

either of these methods.

Ising models with competing ferromagnetic and antiferromagnetic interactions were also the subject of Chapter IV. Here, they appeared as Boltzmann machines, neural networks of stochastic units that could be trained to perform useful tasks. The correlations between adjacent units in these networks were computed by traversing the branches of a tree, in much the same way that quantities in Chapters II and III were computed by summing over paths. A decimation technique, developed originally in the context of the renormalization group, was used to compute these correlations in polynomial time. This led to an efficient learning algorithm for hierarchical Boltzmann machines, one capable of solving non-trivial problems in machine learning and pattern recognition.

During the last decade or so, the computing community has focused mainly on learning in feedforward networks[110]. Boltzmann machines have been somewhat neglected by comparison. The reason for this is easy to understand. In feedforward networks, the activations of units in one layer of the network are deterministic functions of the activations in the previous layer. The operation and training of feedforward networks therefore do not require any stochastic relaxation, only the local propagation of unit activations and error signals. As a result, they are much faster and easier to implement than conventional Boltzmann machines.

The computational tractability of Boltzmann trees opens the door for many future investigations. An important direction of research will be to design Boltzmann trees for new and different types of learning. The problem considered in Chapter IV was one of supervised learning; the network had access to a training set of input

patterns and their target outputs. Many other types of learning can be considered. For example, a weaker form of supervised learning would be to provide only comparative information—that is, whether or not two input patterns have the same output, without specifying the output itself. In many interesting applications, the outputs do not have target values at all; this is the problem of unsupervised learning. A typical problem of this sort is data clustering—separating input patterns into well-defined classes. In this case, the network must extract the underlying structure of the unlabelled input patterns, clearly a much harder problem than its supervised counterpart. Intermediate between supervised and unsupervised learning is the area of reinforcement learning, in which the network receives simple “reward” and “punishment” signals, but does not have access to target values. Work in all these areas is being actively pursued.

This thesis has emphasized the important role of exact computations in the statistical mechanics of disordered systems. Despite considerable progress on certain fronts, clearly, much remains to be understood. The future no doubt has many surprises in store, but of two things, we can be certain. One is that statistical physicists will attempt to tackle problems of greater and greater complexity; the other is that the computers at their disposal will be faster and more powerful. Both these trends hold the promise of many more exciting calculations in the statistical mechanics of disordered systems. This thesis has hopefully pointed in the direction of things to come.

Bibliography

- [1] A. Sommerfeld, *Z. Physik* **47**, 1 (1928).
- [2] P. Weiss, *J. Phys. Radium, Paris* **6**, 667 (1907).
- [3] J. C. Maxwell, *Phil. Mag.* **19**, 19 (1860).
- [4] K. G. Wilson and J. Kogut, *Phys. Rep.* **C12**, 75 (1974).
- [5] S. K. Ma, *Modern Theory of Critical Phenomena* (Addison-Wesley, Redwood City 1976).
- [6] L. D. Landau, *Phys. Z. Sowjetunion* **11**, 26 (1937); V. L. Ginzburg and L. D. Landau, *Zh. Eksp. Teor. Fiz.* **20**, 1064 (1950).
- [7] J. M. Ziman, *Models of Disorder* (Cambridge University Press, New York 1979).
- [8] J. Hertz, A. Krogh, and R. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Redwood City 1991).
- [9] M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific, Singapore 1988).
- [10] J. J. Hopfield and D. W. Tank, *Biological Cybernetics* **52**, 141 (1985).
- [11] S. Kirkpatrick, C. D. Gellatt Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
- [12] D. Amit, *Modelling Brain Function* (Cambridge University Press, Cambridge 1989).
- [13] E. Gardner, *J. Phys. A* **21**, 257 (1988); E. Gardner and B. Derrida, *J. Phys. A* **21**, 271 (1988).
- [14] T. Watkin, A. Rau, and M. Biehl, *Rev. Mod. Phys.* **65**, 499 (1993).
- [15] L. Saul, M. Kardar, and N. Read, *Phys. Rev. A* **45**, 8859 (1991).
- [16] L. Saul and M. Kardar, *Phys. Rev. E* **48**, R3221 (1993).
- [17] L. Saul and M. Jordan, to appear in *Neural Computation* (1994).

- [18] W. Feller, *An Introduction to Probability Theory and Its Applications* vols. 1 & 2 (Wiley, New York 1968).
- [19] C. Itzykson and J.-M. Drouffe, *Statistical Field Theory* (Cambridge University Press, New York 1989).
- [20] J.P. Bouchaud and A. Georges, *Phys. Rep.* **195**, 128 (1990); and references therein.
- [21] E. Ising, *Z. Phys.* **31**, 253 (1925).
- [22] B. M. McCoy and T. T. Wu, *The Two-Dimensional Ising Model* (Harvard University Press, Cambridge, MA 1973).
- [23] M. Kac and J. C. Ward, *Phys. Rev.* **88**, 1332 (1952).
- [24] K. Binder and A. P. Young, *Rev. Mod. Phys.* **58**, 801 (1986).
- [25] K. H. Fischer and J. Hertz, *Spin Glasses* (Cambridge University Press, Cambridge 1990).
- [26] S. F. Edwards and P. W. Anderson, *J. Phys. F* **5**, 965 (1975).
- [27] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations* (MIT Press, Cambridge 1986).
- [28] J. L. McClelland, D.E. Rumelhart, and the PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 2: Psychological and Biological Models* (MIT Press, Cambridge 1986).
- [29] J. J. Hopfield, *Proc. Nat. Acad. Sci., USA* **79**, 2554 (1982).
- [30] E. R. Kandel and J. R. Schwartz, *Principles of Neural Science* (Elsevier, New York 1985).
- [31] T. J. Sejnowski and C. R. Rosenberg, *Complex Systems* **1**, 145 (1987).
- [32] R. P. Lippmann, *Neural Computation* **1**, 1 (1989).
- [33] R. P. Gorman and T. J. Sejnowski, *IEEE Transactions on Acoustics, Speech, and Signal Processing* **36**, 1135 (1988).
- [34] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, *Neural Computation* **1**, 541 (1989).
- [35] D. A. Pomerleau, in *Advances in Neural Information Processing I*, San Mateo: Morgan Kaufmann (1989).
- [36] A. Weigend and N. Gershenfeld, *Time Series Prediction* (Addison-Wesley, Reading, MA 1993).

- [37] F. Rosenblatt, *Principles of Neurodynamics* (Spartan, New York 1962).
- [38] M. L. Minsky and S. A. Papert, *Perceptrons* (MIT Press, Cambridge 1969).
- [39] D. H. Ackley, G.E. Hinton, and T. J. Sejnowski, *Cognitive Science* **9**, 147 (1985).
- [40] A. Ishimaru, *Wave Propagation and Scattering in Random Media* (Academic, New York, 1978).
- [41] *Scattering and Localization of Classical Waves in Random Media*, edited by P. Sheng (World Scientific, Singapore, 1990).
- [42] B. A. Altshuler and P. A. Lee, *Physics Today* **41**, 36 (1988).
- [43] *Chance and Matter*, Les Houches Summer School Proceedings, Session XLVI, edited by J. Souletie, J. Vannimenus, and R. Stora (Elsevier, Amsterdam, 1987).
- [44] M. Kardar and Y.-C. Zhang, *Phys. Rev. Lett.* **58**, 2087 (1987); G. Parisi, *J. Phys. France* **51**, 1595 (1990).
- [45] J.P. Bouchaud, *Europhys. Lett.* **11**, 505 (1990).
- [46] R. Dashen, *J. Math Phys.* **20**, 894 (1979).
- [47] H. de Raedt, A. Lagendijk, and P. de Vries, *Phys. Rev. Lett.* **62**, 47 (1988).
- [48] S. Feng, L. Golubovic, and Y.-C. Zhang, *Phys. Rev. Lett.* **65**, 1028 (1990).
- [49] A.A. Ovchinnikov and N.S. Erikhman, *Sov. Phys. JETP* **40**, 733 (1975).
- [50] A.M. Jayannavar and N. Kumar, *Phys. Rev. Lett.* **48**, 553 (1982).
- [51] L. Golubovic, S. Feng, and F. Zeng, *Phys. Rev. Lett.* **67**, 2115 (1991).
- [52] R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals*, (McGraw-Hill, New York, 1965).
- [53] R. Tao, *Phys. Rev. A* **43**, 5284 (1991).
- [54] E. Medina, M. Kardar, Y. Shapir, and X.-R. Wang, *Phys. Rev. Lett.* **62**, 941 (1989); *Phys. Rev. Lett.* **64**, 15 (1990).
- [55] Y.-C. Zhang, *Phys. Rev. Lett.* **62**, 979 (1989); *Europhys. Lett.* **9**, 113 (1989).
- [56] E. Medina and M. Kardar, *Phys. Rev. Lett.* **66**, 3187 (1991).
- [57] S. John, *Physics Today*, 32 (May 1991).
- [58] E. Medina, M. Kardar, and H. Spohn, *Phys. Rev. Lett.* **66**, 2176 (1991).
- [59] J.P. Bouchaud, D. Touati, and D. Sornette, *Phys. Rev. Lett.* **68**, 1787 (1992).

- [60] P. Sheng and Z. Q. Zhang, *Phys. Rev. B* **48**, 12609 (1993).
- [61] J.T. Chalker and P.D. Coddington, *J. Phys.* **C21**, 2665 (1988); J. T. Chalker and M. Bernhardt, *Phys. Rev. Lett.* **70**, 982 (1993).
- [62] C. Vanneste, P. Sebbah, and D. Sornette, *Europhys. Lett.* **17** (8), 715 (1992).
- [63] D. Chowdbury, *Spin Glasses and Other Frustrated Systems* (Princeton University Press, Princeton, NJ, 1986).
- [64] A. J. Bray and M. A. Moore, *Phys. Rev. Lett.* **58**, 57 (1987).
- [65] D. S. Fisher and D. A. Huse, *Phys. Rev. B* **38**, 386 (1988).
- [66] D. Sherrington and S. Kirkpatrick, *Phys. Rev. Lett.* **35** 192 (1975).
- [67] S. R. McKay, A. N. Berker, and S. Kirkpatrick, *Phys. Rev. Lett.* **48**, 767 (1982).
- [68] G. Toulouse, *Commun. Phys.* **2**, 115 (1977).
- [69] I. Morgenstern and K. Binder, *Phys. Rev. B* **22**, 288 (1980).
- [70] R. R. P. Singh and S. Chakravarty, *Phys. Rev. Lett.* **57**, 245 (1986).
- [71] S. Kirkpatrick, *Phys. Rev. B* **16**, 4630 (1977).
- [72] W. L. McMillan, *Phys. Rev. B* **28**, 5216 (1983).
- [73] R. N. Bhatt and A. P. Young, in *Heidelberg Colloquium on Glassy Dynamics*, edited by J. L. van Hemmen and I. Morgenstern (Springer-Verlag, Heidelberg, 1986).
- [74] J.-S. Wang and R. H. Swendsen, *Phys. Rev. B* **38**, 4840 (1988).
- [75] B. A. Berg and T. Celik, *Phys. Rev. Lett.* **69**, 2292 (1992).
- [76] H. Freund and P. Grassberger, *J. Phys. A* **21**, L801 (1988); H. Freund and P. Grassberger, *J. Phys. A* **22**, 4045 (1989).
- [77] J. A. Blackman and J. Poulter, *Phys. Rev. B* **44**, 4374 (1991); J. A. Blackman, *Phys. Rev. B* **26**, 4987 (1982).
- [78] M. Inoue, *Research Activities* **13**, Sci & Eng., Tokyo Denki Univ., 1 (1991); M. Inoue, private communication.
- [79] H.-F. Cheung and W. L. McMillan, *J. Phys. C.* **16**, 7027 (1983).
- [80] C. Dekker, A. Arts, H. de Wijn, *Phys. Rev. B* **38**, 8985 (1988).
- [81] M. Creutz, *Phys. Rev. B* **43**, 10659 (1991).
- [82] G. Bhanot and J. Lacki, *J. Stat. Phys.* **71**, 259 (1993); **69**, 1941 (1992).

- [83] G. Bhanot, M. Creutz, and J. Lacki, *Phys. Rev. Lett.* **69**, 1841 (1992).
- [84] L. Stodolsky and J. Wosiek, *Nucl. Phys.* **B413**, 813 (1994).
- [85] J. Villain, *J. Phys. C* **10**, 1717 (1977).
- [86] G. Andre, R. Bidaux, J. P. Carton, R. Conte, and L. de Seze, *J. Phys. (Paris)* **40**, 479 (1979).
- [87] G. Forgacs, *Phys. Rev. B* **22**, 4473 (1980).
- [88] R. P. Feynman, *Statistical Mechanics* (Addison-Wesley, Reading, MA, 1972).
- [89] L. Longa and A. M. Oleś, *J. Phys. A* **13**, 1031 (1980).
- [90] M. Kardar, *Phys. Rev. B* **27**, 6869 (1983).
- [91] M. L. Glasser, *Am. J. Phys.* **38**, 1033 (1970).
- [92] H. S. Green and C. A. Hurst, *Order-Disorder Phenomena* (Interscience, London 1964).
- [93] H. Whitney, *Comp. Math.* **4**, 276 (1937).
- [94] R. B. Potts and J. C. Ward, *Progr. Theoret. Phys. (Kyoto)* **13**, 38 (1955).
- [95] S. Wolfram, *Mathematica* (Addison-Wesley, Redwood City, CA 1991).
- [96] M. E. Fisher, *Lectures in Theoretical Physics* (University of Colorado Press, Boulder, 1964), Vol. 7c.
- [97] Y. Abe and S. Katsura, *Prog. Theor. Phys.* **43**, 1402 (1970).
- [98] S. Katsura, *Prog. Theor. Phys.* **38**, 1415 (1967).
- [99] J. Stephenson, *Physica* **148A**, 107 (1988); and references therein.
- [100] A. J. Bray and M. A. Moore, in *Heidelberg Colloquium on Glassy Dynamics*, edited by J. L. van Hemmen and I. Morgenstern (Springer-Verlag, Heidelberg, 1986).
- [101] W. L. McMillan, *J. Phys. C* **17**, 3179 (1984).
- [102] J. C. Angles d'Auriac and R. Maynard, *Solid State Comm.* **49**, 785 (1984).
- [103] F. Barahona, *J. Phys. A* **15**, 3241 (1982).
- [104] F. Calheiros, S. Johannesen, and D. Merlini, *J. Phys. A* **20**, 5991 (1987).
- [105] M. Holzer, *Phys. Rev. Lett.* **64**, 653 (1990).
- [106] C. Itzykson, R. B. Pearson, and J. B. Zuber, *Nucl. Phys.* **B220**, 415 (1983).

- [107] E. Marinari, Nucl. Phys. **B235**, 123 (1984).
- [108] E. Marinari, G. Parisi, and F. Ritort, J. Phys. A **27**, 2687 (1994).
- [109] K. Binder and D. W. Heerman, *Monte Carlo Simulation in Statistical Mechanics* (Springer-Verlag, Berlin 1988).
- [110] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Nature **323**, 533 (1986).
- [111] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge University Press, Cambridge 1986).
- [112] C. Peterson and J. R. Anderson, Complex Systems **1**, 995 (1987).
- [113] G. E. Hinton, Neural Comp. **1**, 143 (1989).
- [114] C. C. Galland, Network **4**, 355-379 (1993).
- [115] J. J. Hopfield, Proc. Nat. Acad. Sci. USA **84**, 8429 (1987).
- [116] E. Yair and A. Gersho, in *Advances in Neural Information Processing Systems I*, ed. D.S. Touretzky (Morgan Kaufmann, San Mateo 1988).
- [117] Y. Freund and D. Haussler, in *Advances in Neural Information Processing Systems* (Morgan Kaufman, San Mateo 1992).
- [118] T. J. Sejnowski, P. K. Kienker, and G. E. Hinton, Physics **22D**, 260 (1986).
- [119] T. P. Eggarter, Phys. Rev. B **9**, 2989 (1974).
- [120] M. F. Møller, Neural Networks **6**, 525 (1993).
- [121] M. Frean, Neural Comp. **2**, 198 (1990).
- [122] J. Pearl, Artificial Intelligence **19**, 241 (1986).
- [123] J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kauffman, San Mateo 1988).
- [124] S. L. Lauritzen and D. J. Spiegelhalter, J. Roy. Stat. Soc. B **50**, 157 (1988).
- [125] D. Cule and Y. Shapir, preprint (1993).
- [126] R. Friedberg and Y. Yu, preprint (1993).

About the Author

Lawrence Kevin Saul was born on August 6, 1969 in Los Angeles, CA, where he attended St. Michael's Elementary School and Harvard High School. He received a National Merit Scholarship to attend Harvard College in 1986. He graduated summa cum laude with a bachelor's degree in physics in 1990 and received an ONR fellowship to pursue his graduate studies at M.I.T. His main interests in graduate school were statistical mechanics and neural computation. He will spend part of next summer at AT&T Bell Laboratories in Murray Hill, NJ, then return to M.I.T. as a postdoctoral fellow in the Department of Brain and Cognitive Sciences.