6

Finite-temperature quantum spin algorithms

The simulation of quantum spin systems at finite temperatures is a very active research area not only because of the importance of the problems being simulated but also because of the availability of highly effective Monte Carlo methods, such as the the loop/cluster and worm algorithms, for many problems of interest. We previewed these algorithms in Chapters 4 and 5. In this chapter and its associated appendices, we provide a more complete and formal derivation and discussion of them. We also discuss the conditions under which simulations at low but finite temperature may be used to study ground-state properties and quantum critical phenomena. While we mainly detail the applications of these algorithms to quantum spin systems, we conclude with a discussion of their use (with minor modifications) to simulate Bosonic lattice models.

6.1 Feynman's path integral

In Section 5.2.5, we used the path-integral formalism to present a simple example of a loop/cluster algorithm for a quantum spin model. Here, we restate the same formalism in a more general and formal language to prepare for a more detailed discussion of loop/cluster updates.

We consider the Hamiltonian

$$H = \sum_{l=1}^{N_l} H_l, \tag{6.1}$$

where H_l is a *local* Hamiltonian defined on some small set of lattice points b_l. A typical local Hamiltonian is $H_l = -J S_{i_l} \cdot S_{j_l}$, where $S_{i_l} = (S_{i_l}^x, S_{i_l}^y, S_{i_l}^z)$ is a spin operator with the standard normalization $S_{i_l} \cdot S_{i_l} = S(S+1)$ in an appropriate basis ($S = 1/2, 1, 3/2, \cdots$). In this case, the local set of lattice points is a pair of nearest-neighbor sites, that is, $b_l = \{i_l, j_l\}$. However, b_l being a set of more than two sites or being just a single site is permissible.

121

In discrete imaginary time, the path integral (5.10) corresponds to the approximation

$$Z = \mathrm{Tr}\,e^{-\beta H} = \mathrm{Tr}\,e^{-\beta \sum_l H_l} \approx \mathrm{Tr} \prod_{k=1}^{M} \prod_{l=1}^{N_l} e^{-\Delta\tau H_l}, \tag{6.2}$$

where M is some positive integer, and $\Delta\tau = \beta/M$. Although the expression on the right-hand side is approximate, its error is proportional to $(\Delta\tau)^2$, and thus vanishes in the limit $M \to \infty$. Because we eventually take this limit to obtain algorithms in continuous imaginary time, the error due to this approximation can be neglected.

We use the product states introduced in (5.5) as our complete orthonormal basis,

$$|\psi\rangle \equiv |s_1\rangle \otimes |s_2\rangle \otimes \cdots \otimes |s_N\rangle,$$

with s_i specifying the local state on the site i. By inserting the identity operator

$$1 = \sum_{\psi=\{s_1,\ldots,s_N\}} |\psi\rangle\langle\psi|$$

between the $e^{-\Delta\tau H_l}$ operators in (6.2), we obtain a representation of the partition function containing only c-numbers,

$$Z \approx \sum_{C=\{\psi(k,l)\}} \prod_{k=1}^{M} \prod_{l=1}^{N_l} \langle \psi(k,l+1)\,|e^{-\Delta\tau H_l}|\,\psi(k,l)\rangle. \tag{6.3}$$

This is a "classicized" representation in terms of the world lines

$$C \equiv \{\psi(1,1), \psi(1,2), \ldots, \psi(M,N_l)\}$$

of spin configurations $\psi(k,l) \equiv \{s_1(k,l), s_2(k,l), \ldots, s_N(k,l)\}$, which satisfy the periodic boundary condition

$$\psi(M+1,l) \equiv \psi(1,l)$$

in imaginary time. We also adopt the notation

$$\psi(k, N_l+1) \equiv \psi(k+1,1).$$

Factorizing (6.3) further, we write

$$\langle \psi(k,l+1)\,|e^{-\Delta\tau H_l}|\,\psi(k,l)\rangle = \left(\prod_{i \notin b_l} \delta_{s_i(k,l+1),s_i(k,l)} \right)$$

$$\times \langle \psi_{b_l}(k,l+1)\,|e^{-\Delta\tau H_l}|\,\psi_{b_l}(k,l)\rangle,$$

where $|\psi_{b_l}\rangle$ is the restriction of $|\psi\rangle$ on the set of lattice points b_l.[1] For example, when $b = \{i,j\}$, $|\psi_b\rangle = |s_i, s_j\rangle$.

---

[1] To be very formal, H_l on the left-hand side is an operator defined on the whole Hilbert space, whereas H_l on the right-hand side is an operator defined on the "local" Hilbert space that concerns b_l only. However, we
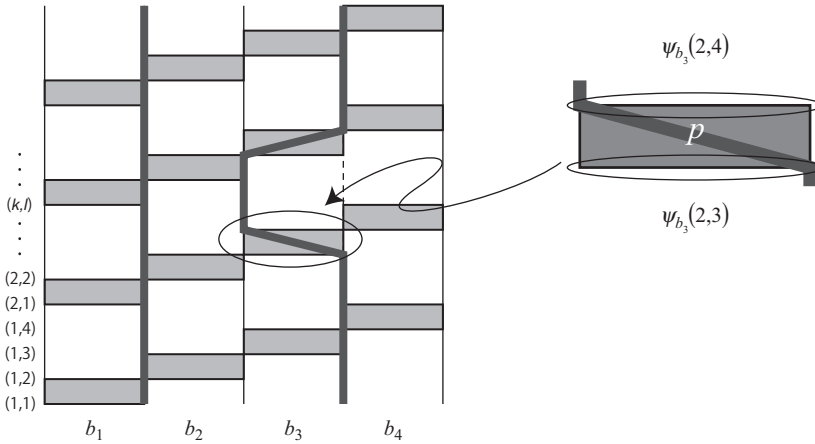
Figure 6.1 The discretized path integral. Each shaded plaquette represents the action of a single bond operator in a time-interval $\Delta\tau$. The N_l consecutive plaquettes for a given k correspond to the imaginary-time evolution operator $\exp(-\Delta\tau H)$.

We call the minimal set of the space-time points involved in a single scattering event a *plaquette*. Namely, a plaquette is a set of points that corresponds to b_l in space and (k, l) and $(k, l + 1)$ in time, for example, a set of four space-time points when b_l represents a pair of sites. With $C_p \equiv \{\psi_{b_l}(k, l+1), \psi_{b_l}(k, l)\}$, where p stands for a plaquette, we can define the weight for the whole space-time configuration as

$$W(C) \equiv \prod_{p=(k,l)} w_p(C_p), \tag{6.4}$$

$$w_p(C_p) \equiv \langle \psi_{b_l}(k, l+1) \left| e^{-\Delta\tau H_l} \right| \psi_{b_l}(k, l) \rangle, \quad [p = (k, l)]. \tag{6.5}$$

Then, our path-integral expression of the partition function becomes

$$Z \approx \sum_{C:\,\Delta(C)=1} W(C), \tag{6.6}$$

with

$$\Delta(C) \equiv \prod_{k,l,i\notin b_l} \delta_{s_i(k,l+1),s_i(k,l)}$$

imposing the constraint on the *world-line configuration C* that the two states $\psi(k, l)$ and $\psi(k, l+1)$, successive in imaginary time, differ only on the plaquettes (Fig. 6.1).

will use the same symbol for both of them here and in what follows as the implication should be obvious from the context.

We note that (6.4), (6.5), and (6.6) correspond to a classical and local expression of the partition function. It is classical in that all the variables $s_i(k, l)$ are c-numbers, and it is local in that all the $w_p(C_p)$ factors are defined on a small set of space-time lattice points.[2]

6.2 Loop/cluster update

6.2.1 General framework

We now discuss the framework of the loop/cluster update algorithm for classical and quantum models. In Section 6.2.3 and Appendices D and E, we present detailed formulas for a few important and illustrative quantum cases. With this framework, we go beyond both the Swendsen-Wang algorithm for the Ising model (Section 4.4) and the quantum loop/cluster update for the $S = \frac{1}{2} XY$ model (Section 5.2.5).

The general framework is most easily understood by analogy with the previously discussed loop/cluster updates. First, let the space of configurations be Σ. On Σ the target weight $W(C)$ is defined and the configuration $C \in \Sigma$ should be generated by it. Now, recalling that the Swendsen-Wang algorithm had graphs allowing only parallel spins and that the loop/cluster update for the XY model had graph elements allowing only certain local world-line configurations, we introduce a set of graphs Γ such that a graph $G \in \Gamma$ is a specifier of a certain set of restrictions on the configuration C. As we did in the Swendsen-Wang algorithm, we define a new weight $W(C, G)$ on the joint phase space $\Sigma \times \Gamma$ such that

$$W(C) = \sum_{G \in \Gamma} W(C, G), \tag{6.7}$$

and construct a Markov chain that samples $W(C, G)$. Equation (6.7) tells us that after marginalizing G, our Markov process will generate a sequence of states C with a frequency proportional to $W(C)$.

Each step of our Markov process has two parts. In the first part, we stochastically generate a graph G for the current configuration C with probability $P(G|C)$, and in the second, we generate a new configuration C' with probability $P(C'|G)$. The conditional probabilities are

$$P(G|C) \equiv \frac{W(C, G)}{W(C)}, \quad P(C|G) \equiv \frac{W(C, G)}{W(G)}, \tag{6.8}$$

---

[2] Here, being "local" means depending only on $O(1)$ number of degrees of freedom. In this sense, long-range interactions can be treated in just the same way as short-range interactions. The problem arising from the large number of interacting pairs can also be handled efficiently (see Appendix E).

and

$$W(G) \equiv \sum_{C \in \Sigma} W(C, G).$$

When we view this combined procedure as one step, the transition probability from the current state C to the next state C' is the sum of the probabilities of all paths leading from C to C',

$$P(C'|C) = \sum_{G \in \Gamma} P(C'|G)P(G|C). \tag{6.9}$$

Using (6.9) and (6.8), we next compute the probability flow from C to C',

$$P(C'|C)W(C) = \sum_{G \in \Gamma} \frac{W(C', G)}{W(G)} \frac{W(C, G)}{W(C)} W(C) = \sum_{G \in \Gamma} \frac{W(C', G)W(C, G)}{W(G)}.$$

Given that the expression on the far right is symmetric with respect to swapping C and C', the detailed balance condition $P(C'|C)W(C) = P(C|C')W(C')$ holds. Therefore, provided the sampling is ergodic, our newly constructed Markov process converges to the correct target distribution $W(C)$. Whether ergodicity holds is something we must examine on a case-by-case basis. Here we just assume that it does.

Viewed as a single Markov step in the joint phase space $\Sigma \times \Gamma$ from (C, G) to (C', G'), the transition probability is simply

$$P\big((C', G') \,|\, (C, G)\big) = P(C'|G')P(G'|C).$$

For this transition probability, the detailed balance condition generally does not hold (Exercise 6.1). However, the Markov process still converges to the target distribution $W(C, G)$. To show this, we use the fact, just proven above, that the Markov process on Σ converges to the target distribution. Therefore, after a sufficiently large number of steps, say, n, $P^{(n)}(C) \propto W(C)$ with an arbitrarily high accuracy. The convergence of the Markov process in the joint phase space immediately follows:

$$P^{(n)}(C, G) = P(C|G)P^{(n)}(G) = P(C|G) \sum_{C'} P(G|C')P^{(n-1)}(C')$$

$$\propto P(C|G) \sum_{C'} P(G|C')W(C') = W(C, G). \tag{6.10}$$

Thus, constructing a loop/cluster update algorithm is equivalent to the problem of finding an appropriate set of graphs Γ and a function $W(C, G)$ that satisfies (6.7). We recall that a local definition of the weight makes the resulting transition procedure particularly easy to use. Therefore, we assume the factorization

$$W(C, G) = \prod_p w_p(C_p, G_p), \tag{6.11}$$

where p represents a small set of lattice sites on which the interaction terms are defined. In classical models, this set is typically a pair of nearest-neighbor spins ($p = \{i, j\}$). In quantum models, it is typically a set of four variables within the path-integral representation. Whereas C stands for the configuration of the whole system, C_p is a part of it, namely, the "local" state on a unit p. The subgraph G_p is defined similarly.

To satisfy (6.7) and (6.11), we now require that

$$w_p(C_p) = \sum_{G_p} w_p(C_p, G_p). \tag{6.12}$$

Because $\sum_G = \prod_p \sum_{G_p}$, (6.7) is satisfied.[3] Further, to update C by a "coin-tossing" process, we require for a fixed G_p that all possible C_p have the same weight, that is, we require

$$w_p(C_p, G_p) = v_p(G_p)\Delta(C_p, G_p), \tag{6.13}$$

where $\Delta(C_p, G_p) = 0, 1$ expresses the local constraints on C_p imposed by the graph element G_p.

With (6.11), (6.12), and (6.13), our transition probability (6.8) finally becomes

$$P(G|C) = \frac{W(C, G)}{W(C)} = \frac{\prod_p w_p(C_p, G_p)}{\prod_p w_p(C_p)} = \prod_p \frac{w_p(C_p, G_p)}{w_p(C_p)}$$

$$= \prod_p P_p(G_p|C_p), \tag{6.14}$$

where

$$P_p(G_p|C_p) \equiv \frac{w_p(C_p, G_p)}{w_p(C_p)} = \Delta(C_p, G_p)\frac{v_p(G_p)}{w_p(C_p)}. \tag{6.15}$$

These equations tell us how to construct a new graph: For each p we are to choose a graph G_p with relative weight $v_p(G_p)$ from the ones compatible with C_p. Then, the whole graph G is simply the union of the local graphs. For the spin configuration update process, we obtain[4]

---

[3] All the examples presented in this book, such as the Swendsen-Wang algorithm and the loop/cluster algorithms for quantum spin models, fit this definition.

[4] This equation can be derived as follows:

$$P(C|G) = \frac{W(C, G)}{W(G)} = \frac{\prod_p w_p(C_p, G_p)}{\prod_p \sum_{C_p} w_p(C_p, G_p)}$$

$$= \frac{\prod_p v_p(G_p)\Delta(C_p, G_p)}{\prod_p \sum_{C_p} v_p(G_p)\Delta(C_p, G_p)} = \frac{\prod_p \Delta(C_p, G_p)}{\prod_p \sum_{C_p} \Delta(C_p, G_p)} = \frac{\Delta(C, G)}{\sum_C \Delta(C, G)}.$$

$$P(C|G) = \frac{\Delta(C,G)}{\sum_C \Delta(C,G)}, \tag{6.16}$$

with

$$\Delta(C,G) \equiv \prod_p \Delta(C_p, G_p).$$

Equation (6.16) means that we can update the state by a simple coin toss, that is, by choosing with equal probability a state from those compatible with the constraints imposed by G. Note that a local degree of freedom s_i may take more than two values in general. In such cases the "coin toss" means choosing among n (> 2) values with equal weight.

6.2.2 Continuous-time loop/cluster update

We now focus on quantum lattice problems, and discuss what algorithm the generic formulas for the transition probabilities (6.14) (with (6.15)) and (6.16) yield in the continuous imaginary-time limit. In doing so, our key task is specifying (6.15).

Let us start with the graphical decompositions (6.12) and (6.13). In the quantum case, the local weight is the matrix element of the local time-evolution operator (6.5). As we did for the $S = \frac{1}{2}$ XY model in Section 5.2.5, we expand the local Hamiltonian H_l as a sum of operators and represent each term in the sum by a graph element,

$$H_l = -\sum_g a_l(g) D_l(g). \tag{6.17}$$

Then, the local imaginary-time evolution operator becomes the operator

$$e^{-\Delta\tau H_l} \approx 1 + \Delta\tau \sum_g a_l(g) D_l(g),$$

whose *matrix elements* are

$$w_p(C_p) \approx \Delta(C_p, g = 0) + \Delta\tau \sum_g a_l(g) \Delta(C_p, g). \tag{6.18}$$

The symbol $\Delta(C_p, g)$ represents the matrix element of $D_l(g)$,

$$\Delta(C_p, g) \equiv \langle \psi_{b_l}(k, l+1) | D_l(g) | \psi_{b_l}(k, l) \rangle.$$

The term with $g = 0$ corresponds to the identity operator,

$$\Delta(C_p, g = 0) = \prod_{i \in b_l} \delta_{s_i(k,l+1), s_i(k,l)},$$

which is represented by the "trivial" graph of vertical segments connecting $s_i(k, l)$ to $s_i(k, l + 1)$. Similarly, the other terms with $\Delta(C_p, g \neq 0)$ are products of Kronecker delta functions that fix the relative values of the local variables bound in g.

By comparing (6.18) with the general formula (6.12) and using (6.13), we identify

$$v_p(g) = \begin{cases} 1 + \mathcal{O}(\Delta\tau) & (g = 0) \\ \Delta\tau \, a_l(g) + \mathcal{O}((\Delta\tau)^2) & (g \neq 0). \end{cases} \tag{6.19}$$

Note that we can neglect the $\mathcal{O}(\Delta\tau)$ term in the $g = 0$ case, since it adds to a constant multiplicative factor to the weight, and can neglect the $\mathcal{O}((\Delta\tau)^2)$ term in the $g \neq 0$ case, because it does not contribute to the weight at all in the $\Delta\tau \to 0$ limit.[5] Then, substituting (6.18) and (6.19) in (6.15), we obtain the graph-assignment probabilities: For C_p not a kink, that is, for the same initial and final state and a $g(\neq 0)$ compatible with C_p,

$$P_p\left(g|C_p\right) = \Delta\tau \, a_l(g) + \mathcal{O}(\Delta\tau^2). \tag{6.20}$$

If C_p is a kink and if g is compatible with C_p,[6]

$$P_p\left(g|C_p\right) = \frac{a_l(g)}{\sum_{g'} a_l(g')\Delta(C_p, g')} + \mathcal{O}(\Delta\tau). \tag{6.21}$$

The first formula (6.20) tells us what to do with the plaquettes with no kinks: To each, we must assign with probability $v_p(g) = \Delta\tau a_l(g)$ a graph element g compatible with the current state. The second formula (6.21) tells us how to update the graph element on plaquettes with kinks: We must choose with a probability proportional to $a_l(g)$ a graph element g among those compatible with the current state and replace the current graph by it.

We now extrapolate to continuous imaginary time. For operations on the kinks, we take the steps described above, that is, we switch the graph elements according to the weight $a_l(g)$. Operations on the nonkinks have a less trivial limit. We regard the k-th imaginary-time layer in the discretized formulation as the one corresponding to the imaginary-time interval $I_k \equiv \{\tau \,|\, k\Delta\tau \leq \tau < (k + 1)\Delta\tau)\}$. In other words, we replace $\psi(k, l)$ by $\psi(\tau)$ with $\tau = k\Delta\tau$. If we consider a finite imaginary-time interval of length I that does not include any kinks, we have $I/\Delta\tau$ plaquettes corresponding to each interaction term H_{b_l}. What we do to each plaquette is assign a graph element of type g with probability $v_p(g) = \Delta\tau a_l(g)$. The number of plaquettes is inversely proportional to $\Delta\tau$, whereas the probability is proportional

---

[5] Because there are only $\mathcal{O}((\Delta\tau)^{-1})$ off-diagonal graphs in the whole system, the $\mathcal{O}((\Delta\tau)^2)$ correction to the off-diagonal term makes an $\mathcal{O}(\Delta\tau)$ contribution that vanishes in the $\Delta\tau \to 0$ limit.

[6] If g is not compatible with C_p, $P_p(g|C_p) = 0$, of course.

Algorithm 17 A Markov step of the quantum loop/cluster algorithm.

Input: A world-line configuration ;

  for each graph element with no kink **do**

    Remove it ;

  end for

  for each kink **do**

    Select a graph element with a probability proportional to $a_l(g)$ from those compatible with the current local state ;

    Replace the current element by it ;

  end for

  for each interaction term l **do**

    for each type of graph element g **do**

      Select imaginary time points $\tau \in [0, \beta)$ with density $a_l(g)$;

      for each selected imaginary time τ **do**

        if the local state is compatible with g **then**

          Place a graph element of type g there ;

        end if

      end for

    end for

  end for

  Connect open ends of neighboring graph elements with vertical lines ;

  Identify loops/clusters ;

  for each loop/cluster **do**

    Flip the loop/cluster (Choose one point on the loop/cluster, select one of its possible local states with equal probability, and assign local states to all the points on the loop/cluster according to the constraint) ;

  end for

  return the updated world-line configuration.

---

to $\Delta\tau$. Therefore, in the $\Delta\tau \to 0$ limit, we obtain a procedure with well-defined probabilities, namely, assigning the type g graphs with density $a_l(g)$. (Everywhere else the trivial graph is to be assigned.) Algorithm 17 summarizes this procedure.

6.2.3 XXZ models

In the previous subsection, we assumed that the Hamiltonian can be decomposed into graphical terms (Eq. (6.17)). While this is not generally true, it is possible for a wide variety of Hamiltonians. An example is the $S = \frac{1}{2}$ XXZ model: $H = \sum_{\langle ij \rangle} H_{ij}$

with $\langle ij \rangle$ denoting nearest neighbors and

$$H_{ij} = -J^{xy}(S_i^x S_j^x + S_i^y S_j^y) - J^z S_i^z S_j^z. \tag{6.22}$$

The most frequently studied $S = \frac{1}{2}$ Hamiltonians, such as the *XY* model and the ferromagnetic and the antiferromagnetic Heisenberg models, are special cases of this model. Even the classical Ising model is a special case of it, and in fact we can show that when the quantum cluster update is applied to (6.22) in the limit $J^{xy} \to 0$, it becomes equivalent to the Swendsen-Wang algorithm for the Ising model (Exercise 6.3). Different anisotropies of the spin interaction, however, yield various types of loops and clusters (Evertz et al., 1993; Evertz and Marcu, 1994; Kawashima, 1996). (In Appendix D we detail the loop/cluster derivations for other important Hamiltonians for which loop/cluster algorithms are possible: the SU(N) model and the SU(N) *J-Q* model.)

When $J^{xy} < 0$, some off-diagonal matrix elements of $e^{-\Delta\tau H_{ij}}$ are negative. If the lattice includes odd cycles, as does the triangular lattice, we cannot avoid the sign problem (Section 5.4). On the other hand, if the lattice lacks odd cycles, as does a bipartite lattice, we can avoid the negative signs by applying the unitary transformation (Marshall transformation, Section 5.2.4) $(S_i^x, S_i^y, S_i^z) \to (-S_i^x, -S_i^y, S_i^z)$ to all sites in one sublattice to change the sign of J^{xy} and thus change the sign of the negative off-diagonal matrix elements. Therefore, we assume $J^{xy} \geq 0$ on bipartite lattices without loss of generality.

The matrix elements of the pair Hamiltonian (6.22) in the basis $\left\{ \left| \frac{1}{2}, \frac{1}{2} \right\rangle, \left| \frac{1}{2}, -\frac{1}{2} \right\rangle, \left| -\frac{1}{2}, \frac{1}{2} \right\rangle, \left| -\frac{1}{2}, -\frac{1}{2} \right\rangle \right\}$ are

$$\left\langle \psi_b' \left| -H_{ij} \right| \psi_b \right\rangle = \begin{pmatrix} \frac{J^z}{4} & 0 & 0 & 0 \\ 0 & -\frac{J^z}{4} & \frac{J^{xy}}{2} & 0 \\ 0 & \frac{J^{xy}}{2} & -\frac{J^z}{4} & 0 \\ 0 & 0 & 0 & \frac{J^z}{4} \end{pmatrix}.$$

We consider four different graph elements (apart from the trivial one corresponding to the identity operator): diagonal (g_{d}), horizontal (g_{h}), binding (g_{b}), and antibinding (g_{ab}). We illustrate these graph elements in Fig. 6.2, together with the matrix elements of the corresponding operators. With these four, the local Hamiltonian can be expressed in three different forms:

$$-H_{ij} = \begin{cases} -\frac{J^z}{4} + \frac{J^{xy}}{2}D_{ij}(g_{\mathrm{d}}) + \frac{J^z - J^{xy}}{2}D_{ij}(g_{\mathrm{b}}) & \text{(I)} \\[2mm] -\frac{J^{xy}}{4} + \frac{J^{xy}+J^z}{4}D_{ij}(g_{\mathrm{d}}) + \frac{J^{xy}-J^z}{4}D_{ij}(g_{\mathrm{h}}) & \text{(II)} \\[2mm] \frac{J^z}{4} + \frac{J^{xy}}{2}D_{ij}(g_{\mathrm{h}}) + \frac{-J^{xy}-J^z}{2}D_{ij}(g_{\mathrm{ab}}) & \text{(III)}. \end{cases} \tag{6.23}$$

Symbol	Graph	$\left\langle s_i' s_j' \left\vert D_{ij}(g) \right\vert s_i s_j \right\rangle$
g_d		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
g_h		$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$
g_b		$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$
g_{ab}		$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

Figure 6.2 The graph elements and the matrix elements of the operators $D_{ij}(g)$ corresponding to each type of graph element. The basis vectors of the two-spin Hilbert space are $\left\vert \frac{1}{2}, \frac{1}{2} \right\rangle$, $\left\vert \frac{1}{2}, -\frac{1}{2} \right\rangle$, $\left\vert -\frac{1}{2}, \frac{1}{2} \right\rangle$, and $\left\vert -\frac{1}{2}, -\frac{1}{2} \right\rangle$. To emphasize the difference in the constraints, we use dashed lines when the connected spins must be anti parallel and solid lines when they must be parallel.

The requirement that $a_l(g)$ be positive selects one of the three:

(I)   easy-axis ferromagnetic model ($0 < J^{xy} \leq J^z$),
(II) easy-plane model ($0 \leq |J^z| < J^{xy}$),
(III) easy-axis antiferromagnetic model ($0 < J^{xy} \leq -J^z$).

In case (I), the clusters consist of the diagonal graph elements g_d, the binding elements g_b, and all the straight vertical lines connecting open ends of these graph elements. A binding graph connects the four spins on a plaquette and forces them to align in the same direction. As a result, all the spins in the same cluster point in the same direction. The resulting graph is not just a collection of loops but is a collection of clusters (that is, loops bound together). The algorithm in case (III) is the same as the one in case (I) except that spins on different sublattices are bound antiparallel while those on the same sublattice are bound parallel. The clusters are simple loops as they are for the *XY* model.

6.2.4 Correlation functions

Why the loop/cluster algorithm often converges much faster than the local update algorithm is best explained by discussing the way we measure the spin-spin correlation functions in the algorithm. We estimate the value of these functions at two space-time points from the frequency with which these points appear in the same loop or cluster. This estimate means that the range of the correlations, the correlation length, corresponds to the typical size of the loops or clusters.

Let us consider a quantity Q expressible as a sum of local operators,

$$Q = \sum_l Q_l.$$

Then

$$\langle Q \rangle \equiv \frac{1}{\beta} \left[\frac{1}{Z(\eta)} \frac{d}{d\eta} Z(\eta) \right]_{\eta \to 0}, \tag{6.24}$$

where

$$Z(\eta) \equiv \mathrm{Tr}\left[e^{-\beta(H - \eta Q)} \right]. \tag{6.25}$$

Accordingly, with (6.5) and the simplified notations $\psi_{b_l}(k, l) \to \psi_p$, $\psi_{b_l}(k, l+1) \to \psi_p'$, $l, b_l \to b$, etc., the weight (6.4) becomes

$$W'(C) \equiv \prod_{p=(k,b)} \left\langle \psi_p' \,\middle|\, 1 - \Delta\tau \left(H_b - \eta Q_b \right) \,\middle|\, \psi_p \right\rangle. \tag{6.26}$$

With the substitution of (6.26) into (6.6), and then (6.6) into (6.24), the thermal average of Q becomes equivalent to the Monte Carlo average of $Q(C)$

$$\langle Q \rangle = \langle Q(C) \rangle_{\mathrm{MC}}, \tag{6.27}$$

where

$$Q(C) \equiv \frac{1}{\beta} \sum_{p=(k,b)} Q(C_p),$$

$$Q(C_p) \equiv \frac{\left\langle \psi_p' \,|\, \Delta\tau Q_b \,|\, \psi_p \right\rangle}{\left\langle \psi_p' \,|\, 1 - \Delta\tau \mathcal{H}_b \,|\, \psi_p \right\rangle} \approx \begin{cases} \Delta\tau \left\langle \psi_p \,|\, Q_b \,|\, \psi_p \right\rangle & (\psi_p' = \psi_p) \\[2mm] \dfrac{\left\langle \psi_p' \,|\, Q_b \,|\, \psi_p \right\rangle}{\left\langle \psi_p' \,|\, (-H_b) \,|\, \psi_p \right\rangle} & (\psi_p' \neq \psi_p). \end{cases} \tag{6.28}$$

After taking the continuous-time limit,

$$Q(C) = \frac{1}{\beta} \int_0^\beta d\tau \, Q(\psi(\tau)) + \frac{1}{\beta} \sum_{(b,\tau):\text{kink}} Q(\psi_b(\tau^+), \psi_b(\tau^-)), \tag{6.29}$$

with

$$Q(\psi(\tau)) \equiv \left\langle \psi(\tau) \left| Q \right| \psi(\tau) \right\rangle \tag{6.30}$$

and

$$Q(\psi_b(\tau^+), \psi_b(\tau^-)) \equiv \frac{\left\langle \psi_b(\tau^+) \left| Q_b \right| \psi_b(\tau^-) \right\rangle}{\left\langle \psi_b(\tau^+) \left| (-H_b) \right| \psi_b(\tau^-) \right\rangle}. \tag{6.31}$$

τ^+ and τ^- are the imaginary times infinitesimally after and before τ. When Q is a diagonal operator, the second term in (6.29) drops. For example, it drops when we estimate the magnetization $Q \equiv \sum_i S_i^z$ in the usual S^z basis. In general, we cannot ignore the second contribution originating from the kinks.

The fact that the loop/cluster algorithm is a Markov process in graph space suggests an alternative estimator defined in terms of graphs rather than states. We recall that in Section 4.4.3, we were able to measure the correlation functions for the susceptibility from the graphs.

To see this in the present case, we write the thermal average of Q as

$$\begin{aligned}
\langle Q \rangle &= \frac{\sum_C W(C) Q(C)}{\sum_C W(C)} = \frac{\sum_{C,G} W(C,G) Q(C)}{\sum_{C,G} W(C,G)} \\
&= \frac{\sum_G W(G) Q(G)}{\sum_G W(G)} = \langle Q(G) \rangle,
\end{aligned}$$

where $Q(G)$ is the fixed-graph average of $Q(C)$ or the graphical estimator of Q:

$$Q(G) \equiv \frac{\sum_C W(C,G) Q(C)}{W(G)}. \tag{6.32}$$

These formulas and an argument similar to the one in Section 4.4.2 yield a specific form of the graphical estimator in such cases. For example, the staggered magnetic susceptibility is defined as

$$\chi_{zz}(q) \equiv \langle Q \rangle, \quad Q \equiv N^{-1} \int_0^\beta d\tau \, M_q^\dagger(\tau) M_q(0),$$

where $M_q \equiv \sum_i e^{iqr_i} S_i^z$ and $M_q(\tau) \equiv e^{\tau H} M_q e^{-\tau H}$. When we set $q = 0$, we obtain the uniform susceptibility. The ordinary estimator of this quantity is

$$Q(C) = (N\beta)^{-1} \left| \int dX \, e^{iqr_i} S^z(X) \right|^2, \tag{6.33}$$

where $X \equiv (i, \tau)$ and the integral $\int dX$ stands for $\sum_i \int d\tau$. The graphical estimator that yields exactly the same mean value is

$$Q(G) = \frac{1}{N\beta} \sum_{c \in G} |M_c|^2 \tag{6.34}$$

with

$$M_c \equiv \int_c dX\, e^{iqr_i} S^z(X),$$

where the integration is over the space-time region covered by a connected cluster c in the graph G. In the case of the easy-plane *XXZ* model, the clusters are loops, and $M_c(q = 0)$ is simply the total length of the loop. The graphical estimator often yields a more accurate estimate than the ordinary estimator.

In some cases, the difference between the two estimators is not just in the accuracy. We note that an expression such as (6.33) is valid only when the operators are diagonal. Suppose, on the contrary, that the operators involved have nonzero off-diagonal matrix elements, but the corresponding matrix elements of the Hamiltonian are zero. Such a case occurs for the *XXZ* model if we measure the transverse susceptibility

$$\chi_{xx} \equiv N^{-1} \int_0^\beta d\tau \langle M^x(\tau) M^x(0) \rangle,$$

where the total magnetization M^x is in the x direction. In a case like this, we cannot estimate the quantity in an "ordinary" way. With the loop algorithm, as we now show, such nondiagonal quantities are measurable (Brower et al., 1998).

With $X = (i, \tau)$ and $Y = (j, \tau')$ specifying two space-time points and $\mathcal{T}$ indicating the time-ordered product, we estimate the expectation value of

$$Q = (N\beta)^{-1} \mathcal{T} \int dXdY S^x(X) S^x(Y)$$

by expressing its thermal average as

$$\chi_{xx} \equiv \langle Q \rangle = \frac{1}{N\beta} \int dXdY \frac{1}{Z} \sum_C{}' W(C)$$
$$\times \langle s_i(\tau^+) | S^x(X) | s_i(\tau^-) \rangle \langle s_j(\tau^+) | S^x(Y) | s_j(\tau^-) \rangle.$$

Note that we have extended the configuration space by including two discontinuous points in the world-line configuration. The prime in the summation indicates that the summation is over all states that have discontinuities in the world lines at X and Y. Such a state is illustrated in Fig. 6.3. It is obvious that this type of configuration never appears in the loop/cluster simulation. Therefore, it is not possible to measure this quantity by a straightforward generalization of the method discussed so far.
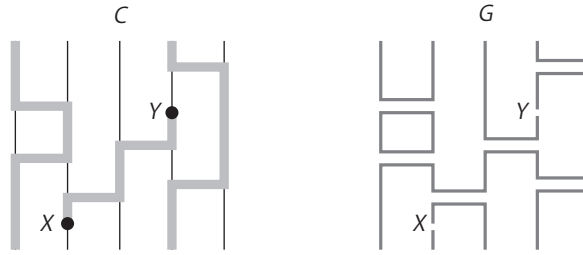
Figure 6.3 Measurement of an off-diagonal correlation function. The world-line configurations (such as "C") and the graphs (such as "G") with discontinuities at two points X and Y contribute.

Therefore, we consider the graphical estimator,

$$\chi_{xx} = \frac{1}{N\beta} \int dX dY \frac{1}{Z} {\sum_{C,G}}' V(G) \Delta(C,G) \frac{1}{4} \delta_{s_i(\tau^+),-s_i(\tau^-)} \delta_{s_j(\tau^+),-s_j(\tau^-)}.$$

First note that in this model there is no way to assign local spin values along a loop if only one discontinuous point lies on the loop. Hence, the summation is zero unless the two points X and Y are on the same loop in G. The result is

$$\chi_{xx} = (N\beta)^{-1} \int dX dY \frac{1}{4} \frac{\sum_G 2^{n_C(G)} V(G) \delta_{c(X),c(Y)}}{\sum_G 2^{n_C(G)} V(G)}$$

$$= \frac{1}{4N\beta} \left\langle \sum_c V_c^2 \right\rangle,$$

where $c(X)$ and $c(Y)$ specify the clusters (loops) to which the points X and Y belong. Thus we obtain a graphical estimator for χ_{xx}:

$$Q(G) \equiv \frac{1}{4N\beta} \sum_{c \in G} V_c^2.$$

The key observation that led to this result was that the summation in the numerator is over the same set of graphs as the summation in the denominator.

The final result for $q = 0$ looks identical to (6.34) even though the quantity discussed is the transverse susceptibility instead of the longitudinal one. This is not a contradiction because M_c and V_c differ by the sign of the spins. For the easy-plane *XXZ* model, the sign of the spins alternates along a loop, and M_c is defined with this alternating sign. V_c is simply the length of the loop. As a result, $|V_c| \geq |M_c|$, which in turn yields $\chi_{xx} \geq \chi_{zz}$. For the easy-axis ferromagnetic *XXZ* model, the sign of the spins does not alternate and M_c is essentially the cluster size. On the other hand, in the above model the graphical estimator of the transverse susceptibility is no longer valid. To make it valid, we replace $\delta_{c(X),c(Y)}$ by a function that is 1 only

when we can split the cluster into two disjoint clusters by cutting it at the points X and Y. In other words, if X and Y are multiply connected in G, the value of the function is 0 even if X and Y are on the same cluster. As a result, χ_{xx} in general becomes smaller than the average cluster size, yielding $\chi_{xx} \leq \chi_{zz}$. Again there is no contradiction. At the isotropic point ($J^{xy} = J^z$), the two estimators are both valid and agree with each other, reflecting the equality of χ_{xx} and χ_{zz} at this point. We note that in both cases (the longitudinal susceptibility in the easy-axis case and the transverse susceptibility in the easy-plane case) the relevant susceptibility corresponds to the average cluster size. Hence, the system update occurs in "units" of clusters whose sizes are roughly the same as the correlation length. This fact is the reason why in these cases the loop/cluster update equilibrates the system very efficiently.

6.2.5 Magnetic fields

None of the quantum spin models studied so far involved an external magnetic field. Given that the frameworks developed in Sections 6.2.1 and 6.2.2 are quite general, we can easily include an external field by simply regarding the field as a one-body interaction that is part of the total interaction.

Let us add the Zeeman term $-H^z S_i^z$ to $S = \frac{1}{2}$ spin models. Then the graphical decomposition is

$$\langle s_i(\tau^+)|H^z S_i^z|s_i(\tau^-)\rangle = H^z \delta_{s_i(\tau^+),\frac{1}{2}} \delta_{s_i(\tau^-),\frac{1}{2}} - \frac{1}{2}H^z \delta_{s_i(\tau^+),s_i(\tau^-)}.$$

We can neglect the second term on the right-hand side as it only shifts the energy by a constant. The first term is a kind of a graph operator. The constraint imposed by this operator or graph element is obviously that the variable on the cluster containing the location (i, τ) is fixed to be $\frac{1}{2}$. According to the general prescription, we are to generate these "fixing" graph elements over the region where spins are aligned with the magnetic field. For the ferromagnetic *XXZ* model, for which the spins in every cluster are aligned with each other, the probability of having at least one fixed graph somewhere in a cluster of volume V is $1 - e^{-H^z V}$, assuming that the cluster is already aligned with the field. This means that we flip the cluster with the probability $\frac{1}{2}e^{-H^z V}$. The factor of $\frac{1}{2}$ comes from the fact that we flip the cluster with probability $\frac{1}{2}$ if there is no fixing element.

We can find the same flipping probability in a different way; namely, instead of taking the Zeeman weight into account in constructing the graphs, we take it into account in the flipping probability. In this approach, we neglect the Zeeman weight when constructing clusters. When we flip them, we do not simply flip them with probability $\frac{1}{2}$, but rather adjust the flipping probability so that the detailed

balance condition is satisfied with respect to the Zeeman weight. When the spins in a cluster of volume V are parallel to the field, flipping the cluster against the field decreases the Zeeman weight from $e^{\frac{1}{2}H^zV}$ to $e^{-\frac{1}{2}H^zV}$. Therefore, if we follow the standard prescription of the Metropolis algorithm, the cluster flipping probability is e^{-H^zV}. Because multiplying the transition probabilities P_{ij} and P_{ji} by the same factor preserves the detailed balance condition, we can make the probability of flipping against the external field $\frac{1}{2}e^{-H^zV}$ and make it $\frac{1}{2}$ for the opposite process. This result is exactly the same as the flipping probability derived in the first approach.

While the two approaches yield the same result in the case of the ferromagnetic *XXZ* model, they do not in general yield the same result when spins are inhomogeneous within the cluster, as is the case for the antiferromagnetic *XXZ* model. For such cases, the second approach is better, although it still might not be good enough to avoid the freezing problem discussed in Section 6.4.1.

6.2.6 Large spins $\left(S > \frac{1}{2}\right)$

The loop algorithm has been discussed just for $S = \frac{1}{2}$ quantum spin models, but the framework presented in Sections 6.2.1 and 6.2.2 is more general. If we were to apply this framework to $S > \frac{1}{2}$ models, each world-line segment in the resulting algorithm would take $2S + 1$ possible values instead of just the two values it takes for the $S = \frac{1}{2}$ case. However, the graphical decomposition is generally not unique, and the efficiency of the resulting algorithm may depend on it. Unfortunately, the existence of a "good" decomposition is not guaranteed. Efficient algorithms for $S > \frac{1}{2}$ spin models that fit the general framework are known only in very few special cases. One example is the SU(N) symmetric model with a fundamental representation that is itself a special case of a $S = (N - 1)/2$ spin model. We discuss this model in Appendix D. Here, we consider an alternative strategy.

The problem can be solved by replacing each spin operator by the sum of $2S$ spins, each carrying $S = \frac{1}{2}$ (Kawashima and Gubernatis, 1994).[7] Let us "split" a spin of magnitude S into $2S$ spins each carrying $S = \frac{1}{2}$:

$$S_i^\alpha \Rightarrow \tilde{S}_i^\alpha \equiv \sum_{\mu=1}^{2S} \sigma_{i\mu}^\alpha \qquad (\alpha = x, y, z), \tag{6.35}$$

where $\sum_{\alpha=x,y,z}(\sigma_{i\mu}^\alpha)^2 = 3/4$. By this replacement, we also expand the Hilbert space so that it now has dimension 2^{2SN}, considerably larger than the original one, $(2S + 1)^N$. To obtain a formulation that exactly corresponds to the original

---

[7] See also Todo and Kato (2001) for an explicit formulation for $S \geq \frac{3}{2}$.

problem, we have to eliminate many states. A projection operator $\hat{P}$ (Kawashima and Gubernatis, 1994; Harada et al., 1998; Todo and Kato, 2001) does this job,

$$Z = \mathrm{Tr}\left[e^{-\beta H(\{S_i\})}\right] = \mathrm{Tr}\left[\hat{P}e^{-\beta H(\{\tilde{S}_i\})}\right].$$

This operator projects the extended Hilbert space onto the original space where all local spin states have the highest S_z value, that is, $(\tilde{S}_i)^2 = S(S+1)$. $\hat{P}$ is the product of the local projection operators $\hat{P}_i$,

$$\hat{P} = \prod_i \hat{P}_i.$$

Because the highest spin state is a symmetric state, each local projection operator is the symmetrizer

$$\hat{P}_i = \frac{1}{(2S)!} \sum_{\pi} D_i(\pi),$$

where the operator $D_i(\pi)$ represents a permutation π that maps $\{1, 2, \ldots, 2S\}$ onto itself. Specifically,

$$D_i(\pi)\left|s_{i,1}, s_{i,2}, \ldots, s_{i,2S}\right\rangle = \left|s_{i,\pi(1)}, s_{i,\pi(2)}, \ldots, s_{i,\pi(2S)}\right\rangle.$$

In terms of the split spins, the pair Hamiltonian (6.22) is now expressed as

$$H_{ij} = \sum_{\mu=1}^{2S} \sum_{\nu=1}^{2S} H_{i\mu,j\nu},$$

where $H_{i\mu,j\nu}$ represents the interaction between the μ-th split spin on the site i and the ν-th split spin on the site j. Each pair Hamiltonian $H_{i\mu,j\nu}$ is simply a Hamiltonian of an $S = \frac{1}{2}$ model. Therefore, we can apply the general prescription discussed previously (Section 6.2.3) directly to the new problem.[8] In the resulting method, $2S$ vertical lines represent each site i (Fig. 6.4). We must repeat the graph-assignment procedure for each of the $(2S)^2$ pairs of indices μ and ν. Apart from this repetition with respect to the split spin indices, the algorithm is identical to the $S = \frac{1}{2}$ algorithm. Specifically, we use exactly the same types of graphs and the same graph-assignment densities as in the $S = \frac{1}{2}$ algorithm.

The remaining task is to cast the projection operator into the framework of the general loop/cluster algorithm. This task is easy as the permutation operator is a type of graph operator. For example, in the case of $S = 1$, the operator $D_i(\pi)$, with π the identity permutation, corresponds to the left diagram at the top of Fig. 6.4 and consists of two vertical world-line segments, whereas if π is the permutation

---

[8]   We have to interpret a pair of sites b as $((i\mu), (j\nu))$ instead of (i, j).
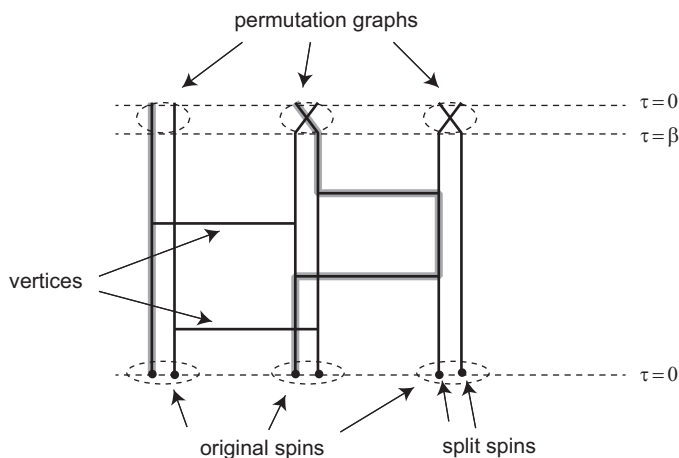
Figure 6.4 A world-line configuration for an $S = 1$ quantum spin system in the split-spin representation.

that swaps the two spins, it corresponds to the middle and the right diagram and consists of two crossing segments. Therefore, as illustrated in Fig. 6.4, inserting the projection operator amounts to picking one of the permutation diagrams compatible with the current state of split spins for each i on the boundaries $\tau = 0, \beta$ and connecting the split spins on i at $\tau = \beta$ to those at $\tau = 0$.

We now briefly mention the calculation of the Haldane gap for various S, as it illustrates the high efficiency of the algorithm. When S is integer, the one-dimensional Heisenberg antiferromagnet is equivalent to the two-dimensional nonlinear σ-model. The latter model is equivalent to the two-dimensional classical Heisenberg model, which does not have an ordered state at any finite temperature. Therefore, for the one-dimensional Heisenberg antiferromagnet with integer spins, we expect a disordered ground state. Accordingly, the excitation gap, which should be proportional to the inverse of the correlation length ξ_τ in the temporal direction, is finite, in contrast to the magnetically ordered state that is gapless due to the Goldstone modes. The gap is called the *Haldane gap* and is measured as the energy difference Δ between the ground and the first excited states. However, since the gap decreases exponentially fast as the spin magnitude increases, it becomes increasingly difficult to obtain accurate estimates of the gap for larger spins. Monte Carlo simulations (Todo and Kato, 2001) yielded $\xi = 6.0153(3), \Delta = 0.41048(6)$ for $S = 1$, $\xi = 49.49(1), \Delta = 0.08917(4)$ for $S = 2$, and $\xi = 637(1), \Delta = 0.01002(3)$ for $S = 3$ for the spatial spin-spin correlation length ξ, defined as $\langle S_i^z S_j^z \rangle \propto e^{-r_{ij}/\xi}$, and the excitation gap Δ.

6.3 High-temperature series expansion

In Section 6.1, we discussed the mapping of quantum problems to classical problems based on the Feynman path integral. In this section, we derive another mapping based on a high-temperature series expansion. The final results are equivalent. Introducing this second alternative provides a good starting point for discussing the worm update in the next section.

Let us start from the high-temperature series expansion of the partition function truncated at the M-th order:

$$Z_M \equiv \sum_{n=0}^{M} \frac{\beta^n}{n!} \mathrm{Tr}[(-H)^n].$$

If the dimension of the Hilbert space is finite, the series converges absolutely, and when M is increased beyond a certain finite value, it converges exponentially fast. The value of M required for a good approximation is usually of the order of $\beta\Lambda$ where Λ is the typical magnitude of the matrix elements of $-H$.

The first step in the mapping is the introduction of M "boxes," each of which can "contain" the operator $-H$. For each box, we define a variable γ_k that specifies the state of the box: $\gamma_k = 0$, if it is empty, and $\gamma_k = 1$, if it is occupied. Next, we let a state with n filled boxes represent an n-th order term in the series expansion. Since there are $M!/((M-n)!n!)$ ways of filling n boxes, we use the reciprocal of this number as the weight of each such state. Also, because the Hamiltonian is a sum of local interactions, $H = \sum_l H_l$, we partition each box into N_l "cells." The l-th cell might or might not contain H_l. Specifying the empty or filled state of the l-th cell of the k-th box by $\gamma_{kl} = 0$ or 1, we rewrite the partition function as

$$Z_M = \sideset{}{'}\sum_{\{\gamma_{kl}\}} \frac{(M-n)!}{M!} \beta^n \, \mathrm{Tr}\left[\prod_{k=1}^{M} \prod_{l=1}^{N_l} (-H_l)^{\gamma_{kl}} \right], \tag{6.36}$$

where $n \equiv \sum_{kl} \gamma_{kl}$ and the summation is taken over all filling configurations $\{\gamma_{kl}\}$. The prime indicates the restriction $\sum_l \gamma_{kl} \leq 1$ for all k.

6.3.1 Stochastic series expansion

The *stochastic series-expansion* method is a Monte Carlo algorithm that defines a Markov process in the space of configurations $C \equiv \{\gamma_{kl}\}$ with the weight being the summand of (6.36) (Sandvik and Kurkijärvi, 1991). Historically, the first classical representation used in Monte Carlo simulations of a quantum system was the high-temperature series expansion for the $S = \frac{1}{2}$ isotropic Heisenberg model (Handscomb, 1962b). The method was then fully developed by Sandvik and coworkers (Sandvik and Kurkijärvi, 1991; Sandvik, 1992, 1999).

In Section 4.5, we discussed how to define a Markov process to sample a classical high-temperature series expansion. There, the expansion yielded a representation of the partition function as a sum of closed graphs. A major difference between the previous classical case and the current quantum case is that we must now deal with the noncommutivity of the quantum mechanical operators: The order of the operator products in the individual terms of the series expansion matters. To keep track of the order, we introduce an "order" space, which is naturally discrete. This additional dimension plays a role similar to the imaginary-time dimension in the path-integral formalism of Section 6.1.

There are various ways to construct a Markov sampling of the series expansion. However, since the weight is nonlocal, computing the acceptance probability of a proposed change in γ_{kl} is in general computationally expensive. To reduce this expense, we make the weight local by introducing "spin" variables. The trick is by now a familiar one: We insert the closure relation between adjacent operators, obtaining

$$Z_M = \sum_C \sum_G W(C, G), \tag{6.37}$$

where $C \equiv (\psi(1), \psi(2), \ldots, \psi(M))$, $G \equiv (\gamma_{1,1}, \gamma_{1,2}, \ldots, \gamma_{M,N_l-1}, \gamma_{M,N_l})$, and

$$W(C, G) \equiv \frac{(M - n)!}{M!}\, \beta^n \prod_{k=1}^{M} \prod_{l=1}^{N_l} \left\langle \psi(k + 1) \left| (-H_l)^{\gamma_{kl}} \right| \psi(k) \right\rangle.^9 \tag{6.38}$$

It is evident that the weight is local except for the constraint imposed implicitly on G, that is, $\sum_l \gamma_{kl} \leq 1$ and the factor $(M - n)!$. Fortunately, these sources of nonlocality impose only a relatively low computational cost. The Monte Carlo simulation based on the stochastic series expansion is a Markov process in the product space of C and G, with the target distribution being $W(C, G)$ defined in (6.38).

From this new expression, we now define the procedure for updating γ_{kl} while the "world-line configuration" C is fixed. For each k, if there is no filled cell in the k-th box, we choose the l-th cell and fill it with probability $p_l^{(\text{fill})}$.[10] If there is a filled cell in the k-th box, we empty it with probability $p^{(\text{empty})}$. The filling and emptying probabilities satisfy the detailed balance condition

$$p_l^{(\text{fill})} \left\langle \psi(k + 1) \mid \psi(k) \right\rangle = p^{(\text{empty})} \frac{\beta}{M - n} \left\langle \psi(k + 1) \left| (-H_l) \right| \psi(k) \right\rangle.$$

---

[9] Note here that the notation is simplified. The symbol $-H_l$ should be interpreted as $1 \otimes 1 \otimes \cdots \otimes (-H_l) \otimes \cdots \otimes 1$; that is, it is defined as an operator acting on the whole Hilbert space, not on its local component. For spins not in the local unit b_l, it acts as an identity operator.

[10] We construct a cumulative distribution function and use it to pick l. If $\sum_l p_l^{(\text{fill})} < 1$ there is a nonzero probability of not filling any cell.

If $\psi(k+1) \neq \psi(k)$, $p^{(\text{empty})} = 0$ satisfies the equation, so we don't do anything and move to the next box. If $\psi(k+1) = \psi(k)$, on the other hand, there are nontrivial solutions for $p_l^{(\text{fill})}$ and $p^{(\text{empty})}$. The simplest may be

$$p_l^{(\text{fill})} = \frac{\beta}{M-n} \langle \psi(k) \,|\, (-H_l) \,|\, \psi(k) \rangle, \quad p^{(\text{empty})} = 1. \tag{6.39}$$

This choice works if M is large enough to make $\sum_l p_l^{(\text{fill})}$ less than or equal to one for any combination of $\psi(k)$ and $\psi(k+1)$, that is, if $(M-n)/\beta$ is greater than any off-diagonal matrix element of $-H$. Because the typical value of n realized in a simulation is of the order of $\beta \Lambda$, where Λ is the magnitude of a typical matrix element of $-H$, $M \gg \beta \Lambda$ is a sufficient condition for both the validity of the simple choice of the transition probability and the validity of the truncation. Since $\sum_l p_l^{(\text{fill})} \leq 1$, we might not choose any cell even if the current state has no filled cell. In this case, we simply go to the next box.

   Once we have updated G, we must then update C. We may do this by a local update, that is, by changing the "world-line" configuration "locally." However, since we have the constraint that $\psi(k+1)$ and $\psi(k)$ can differ only at a filled cell, the update cannot be done for each cell or box independently. One way of updating the configuration without violating the constraints is to consider each pair of adjacent filled cells located at the same spatial position and attempt to modify the intermediate spin configuration (Fig. 6.5). We determine the probability of accepting an attempt, as usual, from the detailed balance condition. In the present case
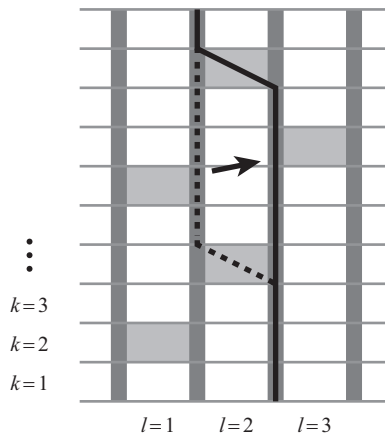

Figure 6.5 Stochastic series expansion with local update. Shaded cells correspond to $\gamma_{kl} = 1$ and white cells are those with $\gamma_{kl} = 0$.

Algorithm 18 Stochastic series expansion with local update.

Input: A world-line (cell) configuration ;
  for each box k **do**
    if $\psi(k+1) \neq \psi(k)$ **then**
      Move to the next box ;
    else
      if there is a filled cell in the box k **then**
        Empty the cell ;
      else
        With probability (6.39), fill the l-th cell ;
      end if
    end if
  end for
  for each filled cell (k, l) **do**
    Find the next filled cell (k', l) ;
    With probability (6.40), change the local states $\psi_l(k), \psi_l(k+1), \ldots, \psi_l(k')$ between the two cells ;
  end for
  return the updated configuration.

$$p_{\text{accept}} = \min(1, \mathcal{R}), \quad \mathcal{R} \equiv \prod_{(k,l)} \frac{\langle \psi'(k+1) \, |(-H_l)| \, \psi'(k)\rangle}{\langle \psi(k+1) \, |(-H_l)| \, \psi(k)\rangle}, \tag{6.40}$$

where $\psi(k)$ and $\psi'(k)$ are the states before and after the modification and the product is over all cells affected by the proposed modification. Algorithm 18 summarizes these procedures.

We can also derive loop/cluster algorithms within the present series-expansion framework. To do so, we simply expand $-H_l$ in terms of graphical operators,

$$-H_l = \sum_g a_l(g) D_l(g).$$

Correspondingly, (6.36) becomes

$$Z_M = \sum_{\{\psi(k)\}} \sum_{\{g_{kl}\}}{}' \frac{(M-n)!}{M!} \beta^n \prod_{k=1}^{M} \left\langle \psi(k+1) \left| \prod_{l=1}^{N_l} a_l(g_{kl}) D_{g_{kl}} \right| \psi(k) \right\rangle,$$

with the convention that $D_l(0) = 1$ and $a_l(0) = 1$. Here, n is the total number of nontrivial graphs, that is, $n \equiv \sum_{kl}(1 - \delta_{g_{kl},0})$. The resulting loop algorithm is described in Algorithm 19.

Algorithm 19 Stochastic series expansion with loop/cluster update.

Input: A world-line (cell) configuration ;
  for each box k **do**
    if $\psi(k+1) \neq \psi(k)$ **then**
      Move to the next box ;
    else
      if there is a filled cell **then**
        Empty the cell ;
      else
        Fill the l-th cell with probability (6.39) ;
        With probability $P(g) = a_l(g)/\sum'_g a_l(g)$, select the type of graph element g to be placed there ;
          ▷ Here, g is chosen among the graph elements that are nontrivial and compatible with the current state on the cell. The summation is over all such graph elements.
      end if
    end if
  end for
  Identify loops/clusters ;
  for each loop or cluster **do**
    Flip the loop/cluster ;
  end for
  return the updated configuration.

6.3.2 *"Continuous-time" limit*

Let us now consider the stochastic series expansion with large M. We will see that we can replace the procedure with infinite M by a statistically equivalent one with a finite number of vertices. In the large-M limit, the discrete index k maps onto an imaginary time $\tau \equiv k\Delta\tau$ and the factor $(M-n)!/M!$ to $1/M^n$, reducing (6.37) to

$$Z_M = \sum_{\{\psi(\tau)\}} \sum_{\{(\tau_v, l_v)\}} (\Delta\tau)^n \prod_v \langle \psi(\tau_v^+) | (-H_{l_v}) | \psi(\tau_v^-) \rangle, \qquad (6.41)$$

where $\Delta\tau = \beta/M$ and $v = (\tau_v, l_v)$ specifies a *vertex* (previously called a "filled cell"). The filling probability (6.39) becomes

$$p_l^{(\text{fill})} = \Delta\tau \langle \psi(k) | (-H_l) | \psi(k) \rangle.$$

We note that the nonlocal nature of the weight due to the factor $M - n$ in (6.39) disappears because $M - n$ is replaced by M. Another source of nonlocality, that is, the constraint of having no more than one filled cell for each k also disappears,

because even if we relax the constraint, the probability of having more than one filled cell is of $\mathcal{O}(\Delta\tau^2)$ and hence is negligible. Therefore, in the infinite M limit, the stochastic series-expansion becomes completely local.

The correspondence between the path-integral and the series-expansion formulations should now be obvious. To obtain the continuous-time algorithm with local update, we simply replace the vertex-assignment procedure in Algorithm 18 with the generation of space-time points, which are uniformly and randomly distributed with a density determined by the diagonal matrix element of $-H_l$, and place interaction vertices at these points. The flipping procedure does not change. We could obtain exactly the same algorithm by starting from the path-integral formulation, but we do not do this here. In the loop/cluster variant, we may again take the infinite M limit with the result being identical to Algorithm 17, which we derived from the path-integral framework.

Since we obtain identical algorithms in the continuous-time limit, the practical value of the series-expansion formulation lies in its application with finite M. Although a finite truncation error then exists, we can neglect it as long as M is much larger than $\beta\Lambda$, because the error is in this case exponentially small. This situation contrasts with that of the path-integral formulation, where the discretization error decreases only slowly (polynomially) as we increase M. Therefore, the stochastic series-expansion method with finite M is the most practical algorithm for many applications. Having a finite and fixed number of boxes is convenient, since we can use a simple array data structure, which makes the programming much simpler than in the case of the continuous-time algorithm where we must use linked-list data structures. On the other hand, the finite M algorithm is difficult to parallelize due to its nonlocal weight, its dependence on the total number of vertices n, and the constraint of having no more than one vertex in a box. These issues make it difficult to express the weight as a simple product of local factors, a fact that in turn makes it difficult to split the task into independent subtasks.

6.4 Worm update

The loop/cluster updates just discussed are very efficient when the loops or clusters created reflect the relevant physical correlations. The natural graphical decomposition of the Hamiltonian for the *XXZ* model discussed in Section 6.2.3 results in such loops or clusters. There are cases, however, in which we are not so lucky. Examples of particular importance are antiferromagnets in a uniform external field and frustrated spin systems. In these cases, the loops or clusters formed in the straightforward application of the method become much larger than the correlation length, and even diverge in size well before the critical point, resulting in the

"freezing" of the simulation. For these cases, an alternative simulation approach based on the worm update can be useful. Although the algorithm is also very useful for dealing with Bosonic systems, we describe the algorithm mainly for spin models in what follows. For applications to a typical Bosonic Hamiltonian, see Section 6.6.

6.4.1 Freezing problem

Let us consider a simple example, a system of only two $S = \frac{1}{2}$ spins coupled antiferromagnetically in a uniform field,

$$H = J\vec{S}_1 \cdot \vec{S}_2 - H^z(S_1^z + S_2^z), \quad J > 0,$$

and see what happens if we apply the loop update to it at low temperature $\beta J, \beta H^z \gg 1$.

First, suppose the current state has both spins pointing up. Since the coupling represented by the horizontal graph in Fig. 6.2 is not compatible with the parallel spins, we cannot assign this graph element. Therefore, the resulting graph consists of only trivial graphs, that is, two vertical loops, each representing a spin. Next, in the loop flipping phase, we attempt to flip each spin. Because the volume of each loop is β, according to the prescription given in Section 6.2.5, the acceptance probability of flipping the loop is thus proportional to $e^{-\beta H^z}$. Since $\beta H^z \gg 1$, for all practical purposes the spins are never flipped and the state remains unchanged.

Now, suppose one of the loops is flipped, against the highly unlikely odds, and the new state has two antiparallel spins. This time, there is nothing that stops the assignment of the horizontal graph elements with density J. The probability of having no graph elements is proportional to $e^{-\beta J}$, and since $\beta J \gg 1$, for all practical purposes we always assign at least one horizontal graph element in this condition. No loop is then a vertical straight line, in contrast to the previous case. As a result, flipping loops never changes the total magnetization or places the system back into the original ferromagnetic state.

It is now clear that the transition between the $S^z = 1$ and $S^z = 0$ states never happens when the temperature is much smaller than J and H^z. The ground state of this system, however, depends on the relative values of J and H^z and changes from the singlet state to the triplet state at $H^z = J$. This observation suggests that in the vicinity of the transition point, the loop update fails to identify the system's ground state.

This example is not an exceptional case, as we encounter similar problems whenever competing interactions, such as antiferromagnetic couplings and a uniform magnetic field, coexist.

6.4.2 Directed-loop algorithm

We can often avoid the freezing problem by using the worm algorithm (Prokov'ev et al., 1998; Kashurnikov et al., 1999). In this section, we present a variant called the *directed loop algorithm* (Sandvik, 1999; Syljuåsen and Sandvik, 2002). While originally used with the series-expansion formulation with finite M, in the following we present it for the infinite M (that is, the continuous imaginary time) framework for conceptual simplicity.

In Section 4.5, we applied the worm update to the high-temperature series expansion of the Ising model. There, the key idea was to break the rule that every site must be shared by an even number of edges. In the worm algorithm, this rule was violated at two sites called *worm head* and *worm tail*. The state was updated by moving the head or tail. In the quantum algorithm described below, we break the rule of conservation of the total S^z at two points in space-time. In other words, world lines terminate at the worm head and tail. Initially, we create the head and tail at the same point or close to each other, and then we move them, leaving a part of a new world line behind or erasing an existing world line. When the head and tail meet, they can annihilate. If they annihilate, the resulting state is the updated spin configuration.

As we discuss below, while the head moves around, it changes its direction of motion only when it is scattered by a *vertex*. After the scattering the head moves in the direction determined in the scattering. Because of this "moment of inertia," the head in the directed-loop algorithm draws a loop more efficiently, avoiding a diffusive "back-and-forth" motion, which is an advantage compared with the usual worm algorithm.

Formally, introducing worms corresponds to adding a source term $\eta \sum_i Q_i$ to the Hamiltonian,

$$H_{\text{worm}} = H - \eta \sum_i Q_i.$$

Here, Q_i corresponds to the head or the tail. In the case of spin models, the typical choice is

$$Q_i = S_i^+ + S_i^-,$$

(in the basis set in which the S_i^z operators are diagonal).[11] In general, Q_i is a Hermitian operator that has only off-diagonal matrix elements. With the source term, the partition function in the series-expansion formulation (6.41) becomes

---

[11] For the Bose-Hubbard model, we would choose $Q_i = b_i^\dagger + b_i$ (in a basis set in which local occupation number operators are diagonal).

$$Z_M = \sum_{\{\psi(\tau)\}} \sum_{\{(\tau_v, l_v)\}} (\Delta\tau)^n \prod_v \langle \psi(\tau_v^+)|(-H_{l_v})|\psi(\tau_v^-)\rangle$$

$$\times \left(1 + \sum_{\substack{(\tau_h, i_h), \\ (\tau_t, i_t)}} (\Delta\tau)^2 \eta^2 \langle \psi(\tau_h^+)|Q_{i_h}|\psi(\tau_h^-)\rangle \langle \psi(\tau_t^+)|Q_{i_t}|\psi(\tau_t^-)\rangle \right), \quad (6.42)$$

where t and h stand for the tail and the head and (τ_t, i_t) and (τ_h, i_h) are the corresponding locations. The summation over $\{\psi(\tau)\}$ is over all configurations satisfying the condition that $\psi(\tau)$ changes only at vertices or at worms. We truncate the series at second order in η since we consider at most one worm head and tail. Including more worms in principle is possible and perhaps useful.

With a slight simplification, the above expression for the partition function becomes

$$Z = \sum_{C,G} W(C, G), \quad (6.43)$$

where $C \equiv \{\psi(\tau)\}$, $G \equiv \{(\tau_v, l_v)\}$ and

$$W(C, G) \equiv (\Delta\tau)^n \prod_v \langle \psi(\tau_v^+)|(-H_{l_v})|\psi(\tau_v^-)\rangle. \quad (6.44)$$

Here, we regard the head and the tail as special vertices and include them in the product over vertices. Correspondingly, the existence or absence of the worm and the positions of its head and tail are now in the definition of G, and the weight of the head and tail is in the definition of H_{l_v}, with l_v being the index of the site on which the head ($v = h$) or tail ($v = t$) lie. Specifically, when G has no worms, $W(C, G)$ is the weight of the stochastic series expansion discussed in Section 6.3.1. If G has a head and a tail, the weight has an extra factor, namely, the second term in the parentheses in (6.42). A "state" in the Markov process of the directed loop algorithm consists of three pieces of information: the spin configuration C, the vertex configuration G, and the direction of the head D.[12] We use the symbol Σ to represent all three, that is, $\Sigma \equiv (C, G, D)$. The directed-loop algorithm is a Markov process in Σ space.

A Markov step consists of four substeps. Starting from a configuration without worms, the first substep is the vertex assignment, the second is the creation of the worm head and tail, and the third is the worm cycle in which the head moves until it comes back to the tail. The fourth and final substep is the pair annihilation of the worm head and tail. The vertex assignment substep proceeds in exactly the same manner as the stochastic series expansion with local update discussed in

---

[12] The weight depends only on C and G, not D.

Section 6.3.2. Namely, we place the vertices over the whole system with density $\langle \psi(\tau)|(-H_l)|\psi(\tau)\rangle$. What is new is the switching between configurations with and without worms and updating the configuration with worms.

Let us consider the creation and annihilation of worms in more detail. We create a worm by first choosing a point uniformly and randomly in space time and then attempting to place the head and tail there. The attempt is accepted with a certain probability. If it is rejected, we count the current state once more as a configuration generated by the Markov process. The inverse process becomes possible when the head returns to the tail. On this occasion, we make an attempt to erase both. The attempt is accepted with a certain probability, and if it is rejected, we reverse the direction of the head and let it keep moving. The configurations before and after the creation of the worm differ only at one space-time point. Due to (6.42), the weight of the worm configuration relative to the wormless one is $(\eta\Delta\tau)^2\langle s_i|Q_i|\sigma\rangle\langle\sigma|Q_i|s_i\rangle$, where σ is the state between the head and the tail, and s_i is the state outside of the pair. Therefore, the detailed balance condition is

$$P_{\text{create}}(s_i, \sigma)\frac{\Delta\tau}{N\beta} = P_{\text{annihilate}}(s_i, \sigma)(\eta\Delta\tau)^2\langle s_i|Q_i|\sigma\rangle\langle\sigma|Q_i|s_i\rangle.$$

$P_{\text{create}}(s_i, \sigma)$ is the probability for creating the head and the tail at the chosen position (i, τ) with the intermediate state being σ and the current local state s_i, and $P_{\text{annihilate}}(s_i, \sigma)$ is the probability of the pair annihilation when the head comes back to the tail. The factor $\Delta\tau/N\beta$ on the left is the probability that the region $\{(i, \tau')|\tau' \in \tau \leq \tau' < \tau + \Delta\tau\}$ is chosen in the creation process. Because η is an artificial field, we can choose it to our advantage by making it sufficiently small so that the annihilation probability is 1. Then

$$P_{\text{create}}(s_i, \sigma) = N\beta\Delta\tau\eta^2\langle s_i|Q_i|\sigma\rangle\langle\sigma|Q_i|s_i\rangle.$$

If we choose $\eta^2 \equiv (\max_{s_i}\sum_\sigma N\beta\Delta\tau\langle s_i|Q_i|\sigma\rangle\langle\sigma|Q_i|s_i\rangle)^{-1}$, we obtain

$$P_{\text{create}}(s_i, \sigma) = \frac{\langle s_i|Q_i|\sigma\rangle\langle\sigma|Q_i|s_i\rangle}{\max_{s_i}\sum_\sigma\langle s_i|Q_i|\sigma\rangle\langle\sigma|Q_i|s_i\rangle}, \quad P_{\text{annihilate}}(s_i, \sigma) = 1. \tag{6.45}$$

In particular, if we take $Q_i \equiv S_i^x$ for the $S = \frac{1}{2}$ *XXZ* model, $P_{\text{create}} = P_{\text{annihilate}} = 1$.

Later we will see that it is convenient to generalize the worm creation and annihilation process. In the above process, we have only two choices: When we have a worm, we either annihilate it or reverse the direction of the head. However, if the local degree of freedom is not a binary variable, when we attempt to create a pair, there are usually multiple choices for the state between the head and the tail (Fig. 6.6). In addition, when the attempted annihilation is rejected and the head turns around, the intermediate state left behind in general can differ from the one before the collision. In order to account for all these possibilities, we have to
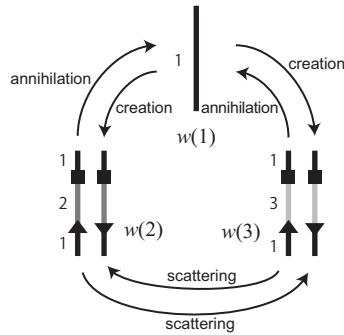
Figure 6.6 The process of creation, annihilation, and scattering of the head (triangle) at the tail (square) when there are three possible local states.

consider the detailed balance condition among a wider class of states. To make the discussion simpler, let us define

$$w_\sigma^t \equiv \begin{cases} (N\beta\eta^2\Delta\tau)^{-1} & (\sigma = s_i) \\ \langle s_i|Q_i|\sigma\rangle\langle\sigma|Q_i|s_i\rangle & (\sigma \neq s_i), \end{cases} \tag{6.46}$$

where the case $\sigma = s_i$ represents the wormless configuration. (The superscript "t" of w^t is for the transition at the "t"ail.) Below we regard the creation/annihilation/scattering process of the head at the tail as the transition among various intermediate states σ. Our task now is to construct the transition probability $p^t(\sigma'|\sigma)$ among the states (or scattering channels) satisfying the condition[13]

$$p^t(\sigma'|\sigma)w_\sigma^t = p^t(\sigma|\sigma')w_{\sigma'}^t.$$

We solve this problem by first finding a solution $w_{\sigma'\sigma}^t$ that satisfies

$$w_{\sigma'\sigma}^t = w_{\sigma\sigma'}^t \geq 0, \quad w_\sigma^t = \sum_{\sigma'} w_{\sigma'\sigma}^t$$

and letting

$$p^t(\sigma'|\sigma) = w_{\sigma\sigma'}^t/w_\sigma^t. \tag{6.47}$$

Then, if $\sigma \neq s_i$, $p^t(\sigma|s_i)$ is the creation probability for the intermediate state σ, and if $\sigma \neq s_i$, $p^t(s_i|\sigma)$ is the annihilation probability when the current intermediate state is σ. In addition, when neither σ nor σ' coincides with s_i, $p^t(\sigma'|\sigma)$ is the probability of turning around and switching the intermediate state to σ'. If we distinguish two states with the same configuration, except for the direction of the head motion, we note that these processes, strictly speaking, do not satisfy the detailed balance

---

[13] Here, p^t in general depends on s_i. This dependence is not explicitly shown.

condition. However, as we discuss in Section 6.4.3, they do satisfy the time-reversal symmetry condition, and this guarantees the convergence of the Markov process to the correct target weight distribution.

Next, we consider updating the worm configurations via a stochastic motion of the head. When we create the head, we assign its initial direction such that it moves away from the tail. After traveling along the vertical line, the head eventually arrives at a vertex. There, it either does or does not change its location and direction of motion. When the vertex represents an m-body interaction, in general $2m$ or more scattering channels exist. A vertex of a two-body interaction, for example, typically has four possibilities for the location of the head after the scattering. We choose stochastically among the four possibilities, taking into account the corresponding weights.

Our problem now is satisfying the time-reversal symmetry condition among the states resulting from the scattering of the head at the vertices. Let us define

$$w_\mu \equiv \left\langle \psi_{b_l}^\mu(\tau_v^+) \left| (-H_l) \right| \psi_{b_l}^\mu(\tau_v^-) \right\rangle \left\langle \psi_{i_h}^\mu(\tau_h^+) \left| Q_i \right| \psi_{i_h}^\mu(\tau_h^-) \right\rangle, \tag{6.48}$$

where $\psi_{b_l}^\mu(\tau_v^+)$ and $\psi_{b_l}^\mu(\tau_v^-)$ are the local states associated with the interaction unit b_l just above and below the temporal position of the vertex τ_v. The label μ specifies the scattering channel. The symbol $\psi_{i_h}^\mu(\tau_h^+)$ represents the local state on the site i_h where the head lies, with τ_h being the temporal position of the head, which is supposed to be close to the vertex. To solve our problem, we follow exactly the same discussion that led to (6.47) for the worm creation and annihilation processes; namely, if we define the mutual weight $w_{\mu\nu}$ such that

$$w_\mu = \sum_\nu w_{\nu\mu}, \text{ and } w_{\mu\nu} = w_{\nu\mu} \geq 0, \tag{6.49}$$

then the scattering probability at vertices, p^v, defined by

$$p^v(\nu|\mu) = w_{\nu\mu}/w_\mu, \tag{6.50}$$

automatically satisfies the time-reversal symmetry condition.

One Monte Carlo step in the directed-loop algorithm is presented in Algorithm 20. The number of cycles N_{cycle} in one step is an arbitrarily fixed number. It is usually set so that every space-time point is visited on average once.[14] For example, in the case of the $S = \frac{1}{2}$ XY model discussed in "Simulation with Worms" in Section 5.2.5, the solution of (6.49) is illustrated in Fig. 6.7. From this solution and (6.50), we obtain the scattering probabilities

$$p^v(1|4) = p^v(2|4) = \tfrac{1}{2}, \quad p^v(4|1) = p^v(4|2) = 1,$$

[14] A sample program that implements this procedure may be found on the Web by searching for the keyword "DSQSS."

Algorithm 20 A Markov step of the directed-loop algorithm.

Input: A world-line configuration ;

  for each vertex with no kink **do**

    Remove it ;

  end for

  for all l **do**

    Pick imaginary times uniform-randomly with density $\langle \psi(\tau)|(-H_l)|\psi(\tau)\rangle$;

    Place vertices corresponding to H_l at these times ;

  end for

  for $i = 1$ to N_{cycle} **do**

    Choose a point in space-time uniformly and randomly :

    With probability $p^{\text{t}}(\sigma|s)$ (6.47), create a head and a tail at this point with the intermediate state being σ ;

    if the attempt is rejected **then**

      Terminate the cycle ;

    else

      Direct the head away from the tail ;

    end if

    loop

      Let the head move until it hits a vertex or the tail ;

      if it hits a vertex **then**

        With probability $p^{\text{v}}(\nu|\mu)$, choose the scattering channel ν (6.50), with μ being the current local state near the vertex ;

        Let the head scatter into the chosen channel ;

      else

        With the probability $p^{\text{t}}(\sigma'|\sigma)$, choose the new intermediate state σ', where σ is the current intermediate state between the head and the tail ;

        if σ' coincides with the state outside of the pair **then**

          Let the head and the tail annihilate ;

          Terminate the cycle ;

        else

          Flip the head's direction and make the intermediate state σ' ;

        end if

      end if

    end loop

  end for

 **return** the updated configuration.

Figure 6.7 The solution of the weight equation (6.49) for the head's scattering at a vertex in the $S = \frac{1}{2}$ XY model. The formula in the lower part of each cell is $w_{\mu\nu}$, defined in (6.49). The shaded entries are the prohibited scatterings. (The coupling constant J^{xy} is denoted as J for simplicity.)

and $p^v(\mu|\nu) = 0$, otherwise. This solution leads to the algorithm described in Section 5.2.5.

6.4.3 Violation of the detailed balance condition

The directed-loop algorithm does not satisfy the detailed balance condition because the head always moves in the current direction and no backward motion is allowed.[15] For example, let C and C' be two subsequent states appearing in this order, in the directed loop simulation. Note that the direction of the head is also a part of the state. The reverse process $C' \rightarrow C$ cannot happen since the head's direction in C' is such that the head moves away from its position in C. This is a clear violation of the detailed balance. Therefore, we need something extra to ensure that the process converges to the correct thermodynamic distribution.

Because the head motion violates the detailed balance condition, we need to prove the stationary condition directly (as in Section 2.6) to show the convergence of the Monte Carlo simulation to the correct stationary distribution. In the present case, the stationary condition reads

$$\sum_{\Sigma} P(\Sigma'|\Sigma)W(\Sigma) = W(\Sigma').$$

---

[15] It is allowed, of course, to change the direction after a scattering event and move in the new direction. What is not allowed is moving against the current direction.

Let us introduce the "time-reversal" operator " ¯ " which is defined such that $\bar{\Sigma}$ is identical to Σ except that the direction of the head is reversed. (If Σ is a wormless state, $\bar{\Sigma} = \Sigma$.) We note first that in Section 6.4.2 we tuned the transition probability such that the "time-reversal symmetry"

$$P(\bar{\Sigma}|\bar{\Sigma}')W(\bar{\Sigma}') = P(\Sigma'|\Sigma)W(\Sigma) \tag{6.51}$$

holds. As is also evident from the description of the directed-loop algorithm in Section 6.4.2, the target weight W is "time-reversal invariant,"

$$W(\bar{\Sigma}) = W(\Sigma). \tag{6.52}$$

From these conditions follows the stationary condition,

$$\sum_{\Sigma} P(\Sigma'|\Sigma)W(\Sigma) = \sum_{\Sigma} P(\bar{\Sigma}|\bar{\Sigma}')W(\bar{\Sigma}') = W(\bar{\Sigma}') = W(\Sigma').$$

Thus, the convergence is guaranteed by the theorem presented in Section 2.6.

As a matter of fact, if we ignore all intermediate states in the directed-loop algorithm, as it transitions from one wormless state to another, it satisfies detailed balance. For example, let us consider two wormless states Σ and Σ' and a sequence of states starting from Σ and ending at Σ': $\Sigma_0, \Sigma_1, \ldots, \Sigma_n$, where $\Sigma_0 = \Sigma$ and $\Sigma_n = \Sigma'$. We can deform the equilibrium probability flowing along this path by using the time-reversal symmetry conditions (6.51) and (6.52).[16] By summing over all intermediate states and the path lengths, we obtain

$$\tilde{P}(\Sigma'|\Sigma)W(\Sigma) = \tilde{P}(\Sigma|\Sigma')W(\Sigma'),$$

where $\tilde{P}$ is the "renormalized" transition probability,

$$\tilde{P}(\Sigma'|\Sigma) \equiv \sum_{n} \sum_{\substack{\Sigma_1,\ldots,\Sigma_{n-1} \\ \text{(with worm)}}} P(\Sigma'|\Sigma_{n-1}) \cdots P(\Sigma_2|\Sigma_1)P(\Sigma_1|\Sigma).$$

6.4.4 Correlation functions

The worm algorithm and the directed-loop algorithm are efficient near a critical point for similar reasons. In these algorithms, an estimator of the correlation

---

[16] In the case of $n = 3$, for example,

$$P(\Sigma'|\Sigma_2)P(\Sigma_2|\Sigma_1)P(\Sigma_1|\Sigma)W(\Sigma) = P(\Sigma'|\Sigma_2)P(\Sigma_2|\Sigma_1)P(\Sigma|\bar{\Sigma}_1)W(\Sigma_1)$$
$$= P(\Sigma'|\Sigma_2)P(\bar{\Sigma}_1|\bar{\Sigma}_2)P(\Sigma|\bar{\Sigma}_1)W(\Sigma_2) = P(\bar{\Sigma}_2|\Sigma')P(\bar{\Sigma}_1|\bar{\Sigma}_2)P(\Sigma|\bar{\Sigma}_1)W(\Sigma')$$
$$= P(\Sigma|\bar{\Sigma}_1)P(\bar{\Sigma}_1|\bar{\Sigma}_2)P(\bar{\Sigma}_2|\Sigma')W(\Sigma'),$$

where Σ and Σ' are wormless configurations and hence $\Sigma = \bar{\Sigma}$ and $\Sigma' = \bar{\Sigma}'$. From this example, it should be clear that the same is true for any n.

functions is the frequency by which the head visits a certain location. For example, to compute the correlation function

$$\Gamma(Y,X) \equiv \langle \mathcal{T} S^x(Y) S^x(X) \rangle$$

where $\mathcal{T}$ indicates the imaginary-time ordered product, we count the number of times the head passes the position $\pm(Y - X)$ relative to the tail. This estimator is quite natural in cases where the loop and directed-loop algorithms are equivalent and the trajectory of the head in the directed-loop algorithm is statistically identical to a loop in the loop algorithm. In what follows, we show that the above estimator is valid even when the directed-loop algorithm is not identical to the loop algorithm.

We start from (6.42) and (6.44) with $Q \equiv S^x$. The correlation function can be expressed as

$$(\Delta\tau\eta)^2 \Gamma(Y,X) = \frac{\sum_{\Sigma' : X,Y} W(\Sigma')}{\sum_{\Sigma' : \text{no worm}} W(\Sigma'),}$$

where $\Sigma = (C, G, D)$ represents a state in the Markov chain and $W(\Sigma)$ is the weight of the worm or wormless state as defined in (6.44). The numerator is the sum over the states with the head at Y and the tail at X or vice versa. Because we encounter the state Σ with a frequency proportional to $W(\Sigma)$ in the Monte Carlo simulation, we can reexpress the correlation function as

$$\Gamma(Y,X) = \frac{1}{\eta^2} \frac{\langle \Delta_{Y,X}(\Sigma) \rangle_{\text{MC}}}{\langle \Delta_\emptyset(\Sigma) \rangle_{\text{MC}}},$$

where $\Delta_{Y,X}(\Sigma)(\Delta\tau)^2 = 1$ if and only if one discontinuity is in the interval $\Delta\tau$ centered at X and the other in the interval $\Delta\tau$ centered at Y. Now, we obtain

$$\Gamma(R) \equiv \frac{1}{N\beta} \int dX dY \Gamma(Y,X) \delta(R - (Y - X))$$

$$= \frac{1}{N\beta\eta^2} \frac{\int dX \langle \Delta_{X+R,X}(\Sigma) \rangle_{\text{MC}}}{\langle \Delta_\emptyset(\Sigma) \rangle_{\text{MC}}}$$

$$= \frac{1}{N\beta\eta^2} \langle n(R) \rangle_{\text{MC}},$$

where $n(R)$ is the average number of times the head passes, during a cycle, the point whose location relative to the tail is R or $-R$.[17]

---

[17] The denominator $\langle \Delta_\emptyset \rangle$ in the last identity is the probability of having no worm at an arbitrarily chosen Monte Carlo step. Thus, it is the inverse of the average length of the trajectory. The numerator is the probability of the relative distance between the head and the tail being R at an arbitrarily chosen Monte Carlo step. Thus, the numerator and the denominator together correspond to the probability of having R at a given step multiplied by the number of steps in one cycle, namely, the average number of times the distance becomes R during one cycle. Hence the last identity.

6.4.5 *XXZ model*

In this section, we explicitly describe the directed-loop update for the case of the $S = \frac{1}{2}$ *XXZ* model. To avoid the explicit dependence on the dimensionality or the lattice geometry, we work with the pair Hamiltonian

$$H_{ij} = -J^{xy}\big(S_i^x S_j^x + S_i^y S_j^y\big) - J^z S_i^z S_j^z - \tfrac{1}{2}h(S_i^z + S_j^z) - E_0 \qquad (6.53)$$

with the exchange coupling $J^{xy} > 0$ and the *pair* magnetic field $h > 0$. The pair magnetic field is the magnetic field per interacting spin pair, for example, $h = H^z/d$ for the d-dimensional hyper-cubic lattice. We have introduced a constant E_0 to ensure that the diagonal elements of H_{ij} are negative. The parameter space of this model divides into six regions, as shown in Fig. 6.8. In solving (6.49), we choose the constant E_0 as $J^z/4$ for region I, $(J^{xy} - h)/4 + h/2$ for regions II_1 and II_2, and $-J^z/4 + h/2$ for regions III_1, III_2, and III_3.

In the special case of the *XY* model without field ($J^z = 0, h = 0$), we obtain the same result as in "Simulation with Worms" in Sections 5.2.6 and 6.4.2. We construct the vertex density and the scattering probability for general S using the one for $S = \frac{1}{2}$ via the coarse-graining prescription discussed below. The results are summarized in Table 6.1. The parameters for $S = \frac{1}{2}$ are found by setting S equal to $\frac{1}{2}$ in this table. Table 6.1 lists the scattering probabilities at the vertices for three scattering directions. The directions of the arrows indicate the direction of the scattering assuming that the head enters the vertex from the lower-left corner. The symbol $p^v(\nearrow | \mu)$, for example, represents the probability for the scattering from the lower left to the upper right, when the current configuration at the vertex is μ. Similarly, the other two symbols $p^v(\downarrow | \mu)$ and $p^v(\rightarrow | \mu)$ represent the probabilities for backward scattering and horizontal scattering.
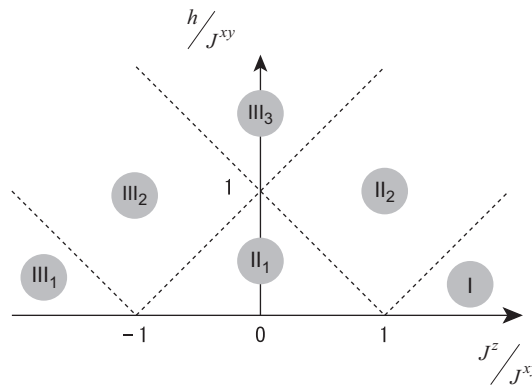

Figure 6.8 The "algorithmic phase diagram" of the *XXZ* model.

Table 6.1 *The vertex density ρ and the scattering probabilities $P(\Gamma|\Sigma)$ of the coarse-grained algorithm for the quantum XXZ spin model with arbitrary S. The "pair magnetic field" h is defined such that for each $S = \frac{1}{2}$ spin the sum of h over all the interacting bonds sharing it is equal to H^z, for example, $h = H^z/(2dS)$ for the d-dimensional hyper cubic lattice. Only the probabilities for nontrivial scattering are shown. (For the initial configurations (μ) not shown in the table, the nontrivial scattering probability is zero, that is, the head always passes straight through the vertex.) $\overline{l} \equiv 2S - l$, $\overline{m} \equiv 2S - m$.*

Case I $(h - J^z \leq -J^{xy})$

$\rho(l, m) = C \equiv \frac{1}{2}[lm(J^z + h) + \overline{l}\,\overline{m}(J^z - h)]$

μ	$p^{\mathrm{v}}(\downarrow|\mu)$	$p^{\mathrm{v}}(\nearrow|\mu)$	$p^{\mathrm{v}}(\rightarrow|\mu)$
$\begin{pmatrix} l & m \\ l_- & m \end{pmatrix}$	$\frac{\overline{m}(-J^{xy}+J^z-h)}{2C}$	$\frac{\overline{m}J^{xy}}{2C}$	0
$\begin{pmatrix} l & m \\ l_+ & m \end{pmatrix}$	$\frac{m(-J^{xy}+J^z+h)}{2C}$	$\frac{mJ^{xy}}{2C}$	0
$\begin{pmatrix} l+1 & m \\ l_- & m+1 \end{pmatrix}$	0	$\frac{1}{l}$	0
$\begin{pmatrix} l-1 & m \\ l_+ & m-1 \end{pmatrix}$	0	$\frac{1}{l}$	0

Case II$_1$ $(h + J^z \leq J^{xy},\ h - J^z \leq J^{xy})$

$\rho(l, m) = A \equiv \frac{1}{4}[lm(J^{xy} + J^z + 3h) + (l\overline{m} + \overline{l}m)(J^{xy} - J^z + h) + \overline{l}\,\overline{m}(J^{xy} + J^z - h)]$

μ	$p^{\mathrm{v}}(\downarrow|\mu)$	$p^{\mathrm{v}}(\nearrow|\mu)$	$p^{\mathrm{v}}(\rightarrow|\mu)$
$\begin{pmatrix} l & m \\ l_- & m \end{pmatrix}$	0	$\frac{m(J^{xy}+J^z-h)}{4A}$	$\frac{m(J^{xy}-J^z-h)}{4A}$
$\begin{pmatrix} l & m \\ l_+ & m \end{pmatrix}$	0	$\frac{m(J^{xy}+J^z+h)}{4A}$	$\frac{\overline{m}(J^{xy}-J^z+h)}{4A}$
$\begin{pmatrix} l+1 & m \\ l_- & m+1 \end{pmatrix}$	0	$\frac{J^{xy}+J^z+h}{\overline{l}\cdot 2J^{xy}}$	$\frac{J^{xy}-J^z-h}{\overline{l}\cdot 2J^{xy}}$
$\begin{pmatrix} l-1 & m \\ l_+ & m-1 \end{pmatrix}$	0	$\frac{J^{xy}+J^z-h}{l\cdot 2J^{xy}}$	$\frac{J^{xy}-J^z+h}{l\cdot 2J^{xy}}$

(*continued*)

Table 6.1 (*cont.*)

Case II$_2$ ($J \leq h + J^z$, $-J^{xy} \leq h - J^z \leq J^{xy}$)

$\rho(l,m) = A \equiv \frac{1}{4}[lm(J^{xy} + J^z + 3h) + (l\overline{m} + \overline{l}m)(J^{xy} - J^z + h) + \overline{l}\,\overline{m}(J^{xy} + J^z - h)]$

μ	$p^{\text{v}}(\downarrow|\mu)$	$p^{\text{v}}(\nearrow|\mu)$	$p^{\text{v}}(\rightarrow|\mu)$
$\begin{pmatrix} l & m \\ l_- & m \end{pmatrix}$	0	$\frac{\overline{m}(J^{xy}+J^z-h)}{4A}$	0
$\begin{pmatrix} l & m \\ l_+ & m \end{pmatrix}$	$\frac{m(-J^{xy}+J^z+h)}{2A}$	$\frac{mJ^{xy}}{2A}$	$\frac{\overline{m}(J^{xy}-J^z+h)}{4A}$
$\begin{pmatrix} l+1 & m \\ l_- & m+1 \end{pmatrix}$	0	$\frac{1}{l}$	0
$\begin{pmatrix} l-1 & m \\ l_+ & m-1 \end{pmatrix}$	0	$\frac{J^{xy}+J^z-h}{l\cdot 2J^{xy}}$	$\frac{J^{xy}-J^z+h}{l\cdot 2J^{xy}}$

Case III$_1$ ($h + J^z \leq -J^{xy}$)

$\rho(l,m) = B \equiv lmh + (l\overline{m} + \overline{l}m)\frac{-J^z+h}{2}$

μ	$p^{\text{v}}(\downarrow|\mu)$	$p^{\text{v}}(\nearrow|\mu)$	$p^{\text{v}}(\rightarrow|\mu)$
$\begin{pmatrix} l & m \\ l_- & m \end{pmatrix}$	$\frac{m(-J^{xy}-J^z-h)}{2B}$	0	$\frac{mJ^{xy}}{2B}$
$\begin{pmatrix} l & m \\ l_+ & m \end{pmatrix}$	$\frac{\overline{m}(-J^{xy}-J^z+h)}{2B}$	0	$\frac{\overline{m}J^{xy}}{2B}$
$\begin{pmatrix} l+1 & m \\ l_- & m+1 \end{pmatrix}$	0	0	$\frac{1}{l}$
$\begin{pmatrix} l-1 & m \\ l_+ & m-1 \end{pmatrix}$	0	0	$\frac{1}{l}$

Case III$_2$ ($-J^{xy} \leq h + J^z \leq J^{xy}$, $J^{xy} \leq h - J^z$)

$\rho(l,m) = B \equiv lmh + (l\overline{m} + \overline{l}m)\frac{-J^z+h}{2}$

μ	$p^{\text{v}}(\downarrow|\mu)$	$p^{\text{v}}(\nearrow|\mu)$	$p^{\text{v}}(\rightarrow|\mu)$
$\begin{pmatrix} l & m \\ l_- & m \end{pmatrix}$	0	0	$\frac{m(J^{xy}-J^z-h)}{4B}$
$\begin{pmatrix} l & m \\ l_+ & m \end{pmatrix}$	$\frac{\overline{m}(-J^{xy}-J^z+h)}{2B}$	$\frac{m(J^{xy}+J^z+h)}{4B}$	$\frac{\overline{m}J^{xy}}{2B}$
$\begin{pmatrix} l+1 & m \\ l_- & m+1 \end{pmatrix}$	0	$\frac{J^{xy}+J^z+h}{\overline{l}\cdot 2J^{xy}}$	$\frac{J^{xy}-J^z-h}{\overline{l}\cdot 2J^{xy}}$
$\begin{pmatrix} l-1 & m \\ l_+ & m-1 \end{pmatrix}$	0	0	$\frac{1}{l}$

Table 6.1   (*cont.*)

Case III$_3$ ($J^{xy} \leq h + J^z$, $J^{xy} \leq h - J^z$)

$\rho(l,m) = B \equiv lmh + (l\overline{m} + \overline{l}m)\frac{-J^z + h}{2}$

μ	$p^{\mathrm{v}}(\downarrow\|\mu)$	$p^{\mathrm{v}}(\nearrow\|\mu)$	$p^{\mathrm{v}}(\rightarrow\|\mu)$
$\begin{pmatrix} l & m \\ l_- & m \end{pmatrix}$	0	0	0
$\begin{pmatrix} l & m \\ l_+ & m \end{pmatrix}$	$\dfrac{\begin{array}{c} m(-J^{xy} + J^z + h) \\ +\overline{m}(-J^{xy} - J^z + h) \end{array}}{2B}$	$\dfrac{mJ^{xy}}{2B}$	$\dfrac{\overline{m}J^{xy}}{2B}$
$\begin{pmatrix} l+1 & m \\ l_- & m+1 \end{pmatrix}$	0	$\dfrac{1}{l}$	0
$\begin{pmatrix} l-1 & m \\ l_+ & m-1 \end{pmatrix}$	0	0	$\dfrac{1}{l}$

An example that illustrates the utility of the directed-loop algorithm is the calculation of the magnetization of the antiferromagnetic Heisenberg chain as a function of the uniform external field. As we discussed in Section 6.4.1, a magnetic field competing with the exchange couplings causes a freezing problem in the loop algorithm, so this method cannot be used to compute a magnetization curve at low temperatures. With the directed-loop algorithm, Syljuåsen and Sandvik (2002) obtained accurate data using a reasonable amount of computer time. However, their results also showed that the autocorrelation time increases as the system size increases in the intermediate regions between successive magnetization plateaus, suggesting that the slowing down is tamed but not completely eliminated even by the worm algorithm. To date, no complete solution to this problem is known.

Now, let us consider the problem of finding the scattering probabilities for general S using the one obtained for $S = \frac{1}{2}$. In principle, we could directly solve (6.49), but the solution is not unique, and the efficiency of the resulting algorithm largely depends on which solution we choose.

Below we present the result of another approach based on the split-spin representation (Section 6.2.6). We can reformulate the model with $S > \frac{1}{2}$ in terms of $2S$ Pauli spins: $S_i \rightarrow \sum_\mu \sigma_{i\mu}$. Doing so leads to an algorithm in which the head moves in a space-time manifold of $2S$ vertical lines for each site. Now let us suppose that we visualize the simulation and make a real-time animation of it. Then, imagine we look at it on a low-resolution monitor. The $2S$ lines will blur and appear as a single thick line. In the blurred image, we cannot tell on which of the $2S$ lines the head μ is. We can tell only on which site i and at what time τ it is located. Similarly, we cannot tell on which of the $2S$ lines a particular vertex is footed, but again we
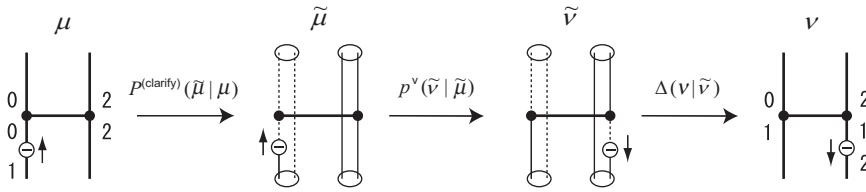
Figure 6.9   The derivation of the scattering probability of the head in the "blurred" algorithm. μ and ν are the initial and the final states in the blurred image, whereas $\tilde{\mu}$ and $\tilde{\nu}$ are the corresponding states in the split-spin image. The numbers associated with the vertical lines in the first and the last diagram represent the "brightness" of the line.

can identify the site and time. Suppose the single line in the blurred image looks brighter when in the original image we have more up-spins than down-spins on the $2S$ lines, that is, we have $2S + 1$ brightness levels (distinguishable on the monitor). The head changes the brightness level of the line by one.

Harada and Kawashima (2002) noted that such a blurred animation can be generated without first creating the sharp image and viewing it on the low-resolution monitor. Instead we could set up our simulation with a set of transition matrices defined directly in terms of the brightness. Note also that for all practical purposes what we need is only the blurred animation. In short, the split-spin representation is not necessary for describing the algorithm or writing computer codes, but as we discuss below, it is useful for the derivation of the algorithm.

It is straightforward to implement the split-spin idea by defining appropriate scattering probabilities. As an example we consider the $S = 1$ antiferromagnetic Heisenberg model. Suppose that the head has just hit a vertex in the state μ (the first diagram in Fig. 6.9). The probability of obtaining the last diagram as the final state of the scattering is

$$p^{\mathrm{v}}(\nu|\mu) = \sum_{\tilde{\mu}\tilde{\nu}} \Delta(\nu|\tilde{\nu})p^{\mathrm{v}}(\tilde{\nu}|\tilde{\mu})P^{(\mathrm{clarify})}(\tilde{\mu}|\mu). \tag{6.54}$$

The probability $P^{(\mathrm{clarify})}(\tilde{\mu}|\mu)$ that the original (sharp) image $\tilde{\mu}$ is associated with the blurred image μ is proportional to the weight of the original image,

$$P^{(\mathrm{clarify})}(\tilde{\mu}|\mu) = \frac{\tilde{w}(\tilde{\mu})\Delta(\mu|\tilde{\mu})}{\sum_{\tilde{\mu}} \tilde{w}(\tilde{\mu})\Delta(\mu|\tilde{\mu})},$$

where $\Delta(\mu|\tilde{\mu}) = 1$ if and only if μ is the blurred image of the original one $\tilde{\mu}$. The weight $\tilde{w}(\tilde{\mu})$ is the weight in the split-spin representation,

$$\tilde{w}(\tilde{\mu}) = \sum_{\alpha\beta} \langle \sigma'_{i\alpha}\sigma'_{j\beta} \,|(-H_{i\alpha,j\beta})|\, \sigma_{i\alpha}\sigma_{j\beta} \rangle.$$

The probability p^v is the scattering probability of the head in the split-spin representation (in the sharp image). Here we have not explicitly introduced the worm weight.[18]

In a similar fashion, we obtain the worm creation/annihilation probabilities from the blurring of the corresponding process in the split-spin representation. In the $S = \frac{1}{2}$ model, if the current local spin state at the chosen point is $+\frac{1}{2}$, we place a head corresponding to the operator S^+ above the one corresponding to S^- (switching the local state between them to $-\frac{1}{2}$). We do the opposite if the local spin state is $-\frac{1}{2}$. When coarse-grained, these placements amount to the following: When the local spin state is l, the probability of creating an S^+ head above an S^- head is $l/2s$; otherwise, we create S^- above S^+. For the worm annihilation, if S^+ is above S^- before the rendezvous and the local spin state outside of the interval is l (the local state between the two is $l - 1$), the two heads are on the same line with the probability l^{-1}, which is the annihilation probability. If, on the other hand, the S^+ is below S^- just before the rendezvous, the probability is $(2S - l)^{-1}$.

Finally, we assign the vertex density as follows: Let us consider an interval in which the local spin state is l on one of the two sites and m on the other. In the original (split-spin) image, we assign vertices with density $\langle \sigma'_{i\mu}, \sigma'_{j\nu} | (-H_{i\mu,j\nu}) | \sigma_{i\mu}, \sigma_{j\nu} \rangle$ between the two vertical lines specified by $(i\mu)$ and $(j\nu)$. Therefore, in the blurred image, we assign vertices with density

$$\rho = \sum_{\alpha\beta} \langle s_{i\alpha}, s_{j\beta} | \left(-H_{i\alpha,j\beta} \right) | s_{i\alpha}, s_{j\beta} \rangle = lm\rho_{++} + l\bar{m}\rho_{+-} + \bar{l}m\rho_{-+} + \bar{l}\bar{m}\rho_{--},$$

where $\rho_{\pm\pm}$ is the vertex density for the $S = \frac{1}{2}$ model with the local spin state $(\pm\frac{1}{2}, \pm\frac{1}{2})$.

The resulting vertex densities and the scattering probabilities are shown in Table 6.1. Note that h is now the magnetic field per Pauli-spin pair, for example, $h = H^z/(2dS)$ for the d-dimensional hypercubic lattice. The offset energies E_0 subtracted from the standard definition of the Hamiltonian are $(J^z - h)S^2 + hS$ for regions I and V, $-J^{xy}S^2 + hS$ for regions II, III, and IV, and $J^{xy}S^2 + h(S - 2S^2)$ for region VI.

6.4.6 On-the-fly vertex generation

Section 6.4.2 describes a worm update method in the framework of the series-expansion classicization in which the vertex assignment and the worm update are two separate procedures. In the present subsection, we describe an alternative way

---

[18] Alet et al. (2005) remarked that the present algorithm can be viewed as a directed-loop algorithm that is a special solution to (6.48) with worm weight $w_w \propto \langle s_{i_h}(\tau_h^+) | S^x_{i_h} | s_{i_h}(\tau_h^-) \rangle$.

of introducing worms in the world-line Monte Carlo simulation that generates inter-action vertices "on the fly." The resulting algorithm is quite similar to the original worm algorithm (Prokov'ev et al., 1998); however, the algorithm we present here is a generalization of the directed-loop algorithm and thus has a directed head motion. The "directed-ness" saves some computational time by preventing unnecessary diffusive motion of the head. Compared with the directed-loop algorithm, the on-the-fly algorithm has the same advantage as the conventional worm algorithm; that is, it needs computer memory only for storing the world lines and not for storing "empty space." This advantage is significant in the case of extremely dilute Bose gases, which are often considered in applications to cold atom systems.

The basic idea is very simple. We essentially do the same thing we did in the continuous-time directed-loop algorithm described in Section 6.4.2, but we apply the vertex generation procedure only in a restricted area near the head (Kato and Kawashima, 2009). More precisely, we consider all the interaction terms H_l involv-ing the head (or the site on which the head is currently located). For the imaginary-time interval I, delimited by the head itself and the kink ahead of it, we attempt to place a vertex for each interaction term following the rule described in Section 6.4.2; that is, we place it by a Poisson process whose rate equals the diagonal matrix element of H_l. Among such vertices, we pick the one closest to the head. As described in Section 6.4.2, we move the head to the vertex and let it scatter in exactly the same fashion as in the conventional directed-loop algorithm. Once the scattering is done, the vertices are erased and we repeat this procedure again and again until the head annihilates with the tail. It is possible that no new vertex is placed between the head's current position and the next kink. In this case, we simply move the head to the kink delimiting I and let it scatter there, again following the same rule as before.

While this procedure gives us correct results, we can improve it by skipping all the "forward-scatterings." In the procedure we just described, at each vertex there is a probability that the head scatters forward with no change in its spatial location or in its direction of motion. Therefore, when this probability is close to unity, we are forced to generate random numbers and make stochastic decisions many times with no change in the world-line configuration until finally some nonforward scattering occurs. To avoid the computational cost of this repetition, we may compute the first scattering time τ_l at which a (nonforward) scattering by the vertex characterized by H_l should take place if no other earlier scattering occurs. After generating τ_l for all interactions, we choose the smallest one, say, τ_{l*}, let the head proceed to τ_{l*}, and scatter at the vertex corresponding to $-H_{l*}$. In this way, we save the computational cost of handling the forward scatterings.

An even better procedure with the same outcome is generating the first scattering time with the "total" scattering rate, that is, the sum of scattering rates of all vertices,

and then decide which is the scattering vertex. This procedure involves a Poisson process only once for each actual nonforward scattering. This task is a little complicated because the state of spins interacting with the current spin varies in imaginary time, which, in turn, makes the decay rate inhomogeneous in the interval I.

There are two simple solutions to this problem. One is to change the definition of I. So far, we have taken I as the interval delimited by the next future kink. Instead, we could define I as the interval delimited by the point at which the environment of the head changes. In other words, we could define I as the maximal interval with a constant "molecular field." Though I in the new definition is shorter than the one previously defined, the new definition serves the current purpose. Now the procedure becomes quite simple. We first generate a uniform random number ζ and let τ_{first} be

$$\tau_{\text{first}} = -\log \zeta / a \tag{6.55}$$

with a being the total scattering amplitude. We then let the head advance by τ_{first}. The total scattering amplitude is the sum of the nonforward scattering amplitudes of all channels, that is,

$$a \equiv \sum_l a_l p_l^{\text{n.f.}}, \tag{6.56}$$

where the summation is taken over all interaction terms in the Hamiltoninan in which the current site is involved, a_l is the matrix element of $-H_l$ evaluated just in front of the head,

$$a_l \equiv \langle \psi(\tau)|(-H_l)|\psi(\tau)\rangle,$$

and $p_l^{\text{n.f.}}$ is the nonforward scattering probability at the vertex representing H_l. To be specific, $p_l^{\text{n.f.}} \equiv 1 - p_l^{\text{v}}(\uparrow | \mu)$, with μ being the local state just in front of the current position of the head.[19] Next, we choose the interaction term l for which the scattering actually occurs with the probability

$$p_l \equiv \frac{a_l p_l^{\text{n.f.}}}{a}.$$

Then, we place a vertex corresponding to H_l just in front of the head and let it scatter there. In this scattering, only nonforward scatterings should be considered since the forward scattering is already taken into account by the fact that we skip the interval τ_{first}. In other words, we should use a modified scattering probability,

$$p^{\text{v}\prime}(\nu|\mu) \equiv \frac{p^{\text{v}}(\nu|\mu)}{p_l^{\text{n.f.}}}, \tag{6.57}$$

---

[19] Here, p^{v} and μ depend on the interaction term l, though the dependence is not explicitly shown.

where $p^v(v|\mu)$ is the previously introduced scattering probability from the initial state μ to the final state ν in the standard procedure (6.50) for the H_l vertex.

Another solution to the problem of a time-dependent scattering rate becomes possible by modifying the above procedure. We replace the (now generally time-dependent) scattering rate a in (6.55) by its maximum value a_{max}, compute

$$\tau_{first} = -\log \zeta / a_{max}, \tag{6.58}$$

and then compensate for the overestimate by rejecting the proposed scattering with a certain probability. To be more precise, a_{max} is any configuration-independent constant larger than the rate in (6.56). In the new prescription, instead of placing a vertex at the time $\tau + \tau_{first}$, we place only a "marker" for it. We move the head to this point and compute the total scattering rate a by (6.56). Then, with the probability $1 - a/a_{max}$, we choose the forward scattering. Otherwise we select the scattering channel with the probability (6.57). This variant is adopted in Algorithm 21.

Because the resulting algorithm for general models looks complicated, it may be helpful to specialize for the simple case of the $S = \frac{1}{2}$ XY model. We first make $w_{worm}(\sigma)$ in (6.46) the same for all cases by adjusting the value of η, which determines the creation and annihilation probabilities of the worm, such that both are unity. The scattering rate for an interaction is always $\frac{1}{4}J^{xy}$ for each bond, making a in (6.58) $\frac{1}{4}zJ^{xy}$ where z is the coordination number. Since a does not depend on the configuration, we take $a_{max} = a$, which avoids having to choose forward scatterings at the marker. The scattering probability of the vertex is unity if there is no kink on it, and the scattering probability at a kink is $\frac{1}{2}$ for diagonal and horizontal scatterings. The resulting algorithm is somewhat simpler than the general one, as shown in Algorithm 22.

6.5 Toward zero temperature

In Part III, we discuss algorithms specifically designed to simulate systems at zero temperature. We now briefly discuss how finite-temperature Monte Carlo methods may be used to study zero-temperature properties. A straightforward way to do this is to run simulations at temperatures that are nonzero but low enough that the behavior is approximately that of the system at zero temperature. This approach is generally successful when there is a finite gap between the ground state and the first excited state. Studying systems with vanishing excitation gap requires a more sophisticated technique. Of particular importance is the application of finite-temperature quantum Monte Carlo simulations to the study of quantum critical phenomena.

Algorithm 21 A cycle in the directed-loop algorithm with "on-the-fly" vertex generation.

Input: A world-line configuration with no worm ;

  Choose a point in space-time uniformly and randomly ;

  Choose σ with probability $p^{\text{t}}(\sigma|s)$ (s is the current state) (6.47) ;

  if $\sigma = s$ **then**

    Return ;

  else

    Create a head and a tail with the intermediate state being σ ;

    Direct the head away from the tail ;

  end if

  loop

    Generate τ_{first} by (6.58) ;

    Place a "marker" at $\tau + \tau_{\text{first}}$;

    Move the head to the position of the first object it encounters ;

    if the next object is the marker **then**

      Compute a defined by (6.56) at the new position ;

      Choose with probability $1 - a/a_{\text{max}}$ whether a forward scattering occurs ;

      if forward scattering **then**

        Return to the beginning of the loop ;

      else

        Choose the interaction term l with probability $p_l \equiv a_l p_l^{\text{n.f.}}/a$;

        Place the vertex corresponding to H_l just in front of the head ;

      end if

      With probability $p^{\text{v}\prime}(\nu|\mu)$ (6.57), let the head scatter into channel ν ;

    else if the next object is a kink **then**

      With probability $p^{\text{v}}(\nu|\mu)$ (6.50), let the head scatter into channel ν ;

    else if the next object is the tail **then**

      Generate σ' with $p^{\text{t}}(\sigma'|\sigma)$ (6.47), where σ is the current intermediate state between the head and the tail ;

      if σ' coincides with the state just behind the head **then**

        Let the head and the tail annihilate ;

        Terminate the loop ;

      else

        Reverse the head's direction and make the intermediate state σ' ;

      end if

    end if

  end loop

  return the updated world-line configuration without worm.

Algorithm 22 Algorithm 21 specialized for the $S = \frac{1}{2}$ *XY* model.

Input: A world-line configuration with no worm ;

  Choose a point in space-time uniformly and randomly ;

  Create a head and tail ;

  Direct the head away from the tail ;

  loop

    Generate a random number ζ ;

    $\tau_{\text{first}} \leftarrow -\log \zeta / (\frac{1}{4} z J^{xy})$;

    Place the marker at the distance τ_{first} from the current position ;

    Let the head proceed until it encounters the marker, a kink, or the tail ;

    if it encounters the marker **then**

      With equal probability, choose a site j from the nearest neighbors to the current site i ;

      Place the vertex corresponding to H_{ij} just in front of the head ;

      if the spins on i and j are parallel **then**

        Let the head scatter diagonally ;

      else

        Let the head scatter horizontally ;

      end if

    else if it encounters a kink **then**

      With probability $\frac{1}{2}$, let the head scatter diagonally or horizontally ;

    else if it encounters the tail **then**

      Let the head and the tail annihilate ;

      Terminate the loop ;

    end if

  end loop

  return the updated configuration.

6.5.1 *Extrapolation to zero temperature*

A system that has a finite-dimensional Hilbert space always has a finite difference ΔE between the ground state energy and the energy of the first excited states. In this case, the expectation value of an arbitrary quantity Q must have a temperature dependence that varies as

$$\langle Q \rangle \sim Q_0 + \mathcal{O}(e^{-\Delta E/kT}), \tag{6.59}$$

in the temperature region $0 < kT \ll \Delta E$. Here, Q_0 is the expectation value of Q at zero temperature. Therefore, if the gap is not too small, we can estimate this expectation value by doing a quantum Monte Carlo simulation with one of the

algorithms discussed in this chapter. We simply need to reach a β much larger than $(\Delta E)^{-1}$. If this is possible, we can regard the results as zero-temperature results. Even better is to run simulations at different temperatures and extrapolate the outcome to zero temperature by using (6.59).

The magnitude of the gap ΔE, however, is usually unknown beforehand. It depends on the nature of the system and might converge to zero in the thermo-dynamic limit $L \to \infty$. A convergence to zero may be related to the occurrence of a spontaneous symmetry breaking. If the order parameter that characterizes the symmetry breaking commutes with the Hamiltonian, the ground state is degenerate even if the system size is finite. A simple example for this is the ferromagnetic, Ising-like *XXZ* model ($0 < |J^{xy}| < J^z$) for which the ground state is either the "all-up" or "all-down" state. The order parameter is the uniform magnetization $M \equiv \sum S_i^z$ and it commutes with the Hamiltonian. Hence the degeneracy is exact. In this case, if we apply the above approach, the energy gap is the difference between the twofold degenerate ground state and the first excited state. Since this gap is $\mathcal{O}(J^z - J^{xy})$, the simulation with $kT \ll \Delta E$ is reasonably straightforward. Therefore, obtaining zero-temperature properties from a simulation at low, but finite temperature is possible.

A more nontrivial example is the easy-axis, antiferromagnetic *XXZ* model on a bipartite lattice. The ground state is the Néel state, and therefore it is twofold degenerate in the thermodynamic limit. The order parameter is the staggered mag-netization $M_s \equiv \sum \epsilon_i S_i^z$ with $\epsilon_i = 1$ for i in one sublattice, and -1 for i in the other. This order parameter does not commute with the Hamiltonian unless $J^{xy} = 0$, which we assume is not the case. As a result the ground state need not be degenerate, and, in fact, it is nondegenerate for any finite system with an even number of spins. We write the ground state of a finite system *symbolically* as $|+\rangle \equiv (|\cdots \uparrow\downarrow\uparrow\downarrow \cdots\rangle + |\cdots \downarrow\uparrow\downarrow\uparrow \cdots\rangle)/\sqrt{2}$, where $|\cdots \uparrow\downarrow\uparrow\downarrow \cdots\rangle$ is a state with positive staggered magnetization (not a perfect, fully polarized, Néel state as the equation may erroneously suggest) and $|\cdots \downarrow\uparrow\downarrow\uparrow \cdots\rangle$ is the degenerate state obtained from the other one by inverting all spins. The first excited state is then $|-\rangle \equiv (|\cdots \uparrow\downarrow\uparrow\downarrow \cdots\rangle - |\cdots \downarrow\uparrow\downarrow\uparrow \cdots\rangle)/\sqrt{2}$, with an exponentially small excitation energy $\Delta E \propto |J^{xy}/J^z|^N$. To single out the ground state, which is unique in a finite system, the temperature must be lower than this exponentially small energy. However, we often do not need to reach such a low temperature since most quantities of interest, such as the energy and the susceptibility, do not depend on which of the two states the system is in. Therefore, the observed values do not strongly depend on the temperature as long as the temperature is below the second excitation level.

Even more subtle cases arise from the spontaneous breaking of a continuous symmetry. In this case, Goldstone's theorem (Nambu, 1960; Goldstone, 1961)

predicts the existence of quasi-particles, Nambu-Goldstone Bosons, that represent long-wave length excitations. The excitation energy of a Nambu-Goldstone Boson is gapless and typically has an algebraic wave-number dependence, that is, $\epsilon_k \propto k^\mu$ where μ is a positive integer. For example, in two or more dimensions the $O(3)$ symmetry is spontaneously broken in the ferromagnetic Heisenberg model. The dispersion of the magnon excitation is $\epsilon_k \sim Ja^2 k^2$, where a is the lattice constant. In a finite system with linear size L, the minimum nonzero wave number is $k_1 \equiv \pi/L$, making the first-excitation gap $\Delta E \sim J(\pi a/L)^2$. Therefore, to probe zero temperature, we need to push the simulation to a temperature low enough so that $\beta J \gg J/(\Delta E) \sim (L/a)^2$ to obtain the ground state. For the antiferromagnetic Heisenberg model, the magnon excitation is $\epsilon_k \sim Jak$. Accordingly, the condition for an effective "zero-temperature" simulation is $\beta J \gg L/a$.

6.5.2 Quantum phase transitions

Many-body problems exhibit phenomena that we cannot predict from the behavior of the individual degrees of freedom. However, when we focus on critical phenomena, we often find behavior that is independent of the details of the interactions. The ability to use simulations to probe this universality is what makes these studies so valuable. Realistic systems contain many different kinds of interactions, and their simulation looks like a formidable problem. A common strategy is to start from a simplified model, identify the critical phenomena, and then consider the effect of various correction terms.

In quantum systems, critical phenomena are just as important and often even more interesting than critical phenomena in classical systems. In the limit of low temperature, it is possible to observe critical phenomena that are specific to quantum systems. We can motivate this fact from the path-integral representation, which maps a d-dimensional quantum system to a $(d + 1)$-dimensional classical system. However, we cannot study zero-temperature quantum critical phenomena using the strategy discussed in the previous subsection, since the system generally has a gapless excitation at the quantum critical point and thus lacks a finite energy scale to mark the entry into the "zero-temperature region." Instead, near the quantum critical point, the gap depends on the system size algebraically, resulting in a slow decrease of the finite-temperature error as a function of system size.

Let us consider physical properties of the one-dimensional transverse-field Ising model, for which the computational aspect was already discussed in Section 5.2.2. Equation (5.6) provides a mapping onto an anisotropic two-dimensional classical Ising model. The quantum model still has a symmetry with respect to the 180-degree rotation around the x-axis. This symmetry is called a Z_2 *symmetry*. At zero temperature and $H^x = 0$, the quantum model exhibits long-range order, that is, the

Z_2 symmetry is spontaneously broken. While the ground state at $H^x = 0$ is fully polarized in the $+z$ or $-z$ direction, as soon as we introduce a finite transverse magnetic field, this state ceases to be an eigenstate of the Hamiltonian. *Quantum fluctuations* generated by the noncommutativity of the first and the second terms in the Hamiltonian cause this change. It is not obvious whether these quantum fluctuations destroy long-range order or merely weaken it. As we discussed in Section 5.2.2, the transverse-field Ising model eventually undergoes a transition from the ordered phase to a disordered phase, as we increase the transverse field.

In Section 5.2.2, we mapped the transverse-field Ising model in one dimension to the two-dimensional classical Ising model with anisotropic couplings (5.19)

$$K^x = \Delta\tau J^z, \quad K^\tau = -2\ln\frac{\Delta\tau H^x}{2}.$$

This mapping connects the critical behavior of the two-dimensional classical anisotropic Ising model (Baxter, 1982) with that of the one-dimensional quantum model. In particular, it establishes that the one-dimensional quantum model is critical when

$$\sinh\frac{K^x}{2}\sinh\frac{K^\tau}{2} = 1. \tag{6.60}$$

By substituting the expressions for K^x and K^τ quoted above in (6.60) and taking the limit $\Delta\tau \to 0$, where the classical mapping becomes exact, we obtain the exact value of the critical field

$$H_c^x = \frac{1}{2}J^z.$$

However, this statement is true only when the thermodynamic limit is properly taken, and to do so we have to be careful, because the thermodynamic limit of the classical system does not necessarily correspond to the thermodynamic limit of the quantum system. To be more specific, if we keep β fixed, the thermodynamic limit of the original one-dimensional quantum model corresponds to the two-dimensional classical model in the limit of infinite aspect ratio in which the size L in the spatial direction is taken to infinity, while the size in the temporal direction, which is β, is kept finite. It is only in the zero-temperature limit of the quantum system that the size of the equivalent classical system becomes infinite in both directions. Therefore, in the present example, the classical mapping guarantees the existence of a critical point only for the zero-temperature phase transition.

In general, when we increase the size of the temporal direction, we observe a crossover from d-dimensional behavior to $(d + 1)$-dimensional behavior. The crossover takes place when the correlation length in the temporal direction in

the $(d + 1)$-dimensional system becomes comparable to the system size in the temporal direction, that is,

$$\xi_t \propto |H^x - H_c^x|^{-z\nu} \sim \beta,$$

where ν is the critical exponent that characterizes the divergence of the correlation length around the quantum critical point, and z is the exponent that relates the length scale to the time scale, called the *dynamical critical exponent*. In the case of the transverse-field Ising model $z = 1$. If $\xi_t \ll \beta$, the system's behavior is identical to the $(d + 1)$-dimensional system since the correlation length is as long as it would be in the infinite-size system. On the other hand, if $\xi_t \gg \beta$, the variation along the temporal direction becomes zero so that the system behaves like a d-dimensional one. This general remark also applies to the transverse-field Ising model, and it follows that this model behaves as a d-dimensional one near the quantum critical point for

$$T > T^*(H^x) \propto |H^x - H_c^x|^{z\nu}.$$

If $d \geq 2$, the d-dimensional classical system itself has a phase transition, and the correlation length diverges near the transition temperature. This finite-temperature phase transition belongs to the d-dimensional Ising universality class. In this case, the crossover from the $(d + 1)$-dimensional behavior to the d-dimensional behavior becomes a genuine phase transition, resulting in the phase diagram schematically sketched in Fig. 6.10.

In the case of the transverse-field Ising model, it is apparent from the mapping that the imaginary-time direction is equivalent to the spatial direction apart from some scaling factor, referred to as the "light (or sound) velocity." This is why $z = 1$ for this model. However, this equivalence does not hold in general. There are cases in which the dimension of the temporal axis is not the same as that of the spatial one.

6.5.3 Finite-size scaling

Knowledge of the renormalization group is indispensable for understanding critical phenomena. Very important for the numerical study of quantum systems is that this formalism enables a finite-size scaling analysis of a zero-temperature critical point. Using finite-size scaling, we are able to characterize a zero-temperature critical point using numerical simulations at finite temperatures. There are many textbooks on this subject (e.g., Ma, 1985; Cardy, 1996; Nishimori and Ortiz, 2010) so we only briefly outline the method.

Let the parameters characterizing the thermodynamic state be $t, h, g, \ldots$, and the system size be L. In the case of the Ising model, for example, t is the deviation of
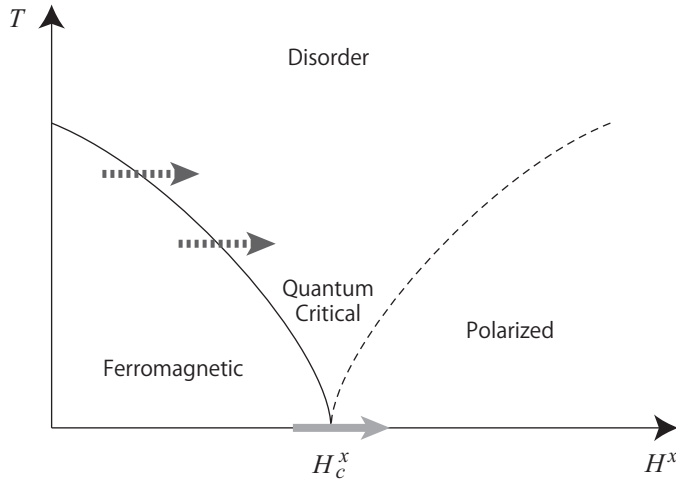
Figure 6.10 Schematic phase diagram of the transverse-field Ising model. The finite-temperature phase transition exists only in two or higher dimensions. When the phase boundary is crossed at nonzero temperature (dotted arrow), the critical phenomena belong to the d-dimensional classical Ising universality class. When it is crossed at zero temperature (solid arrow), it belongs to the $(d+1)$-dimensional classical Ising universality class. In one dimension, there is a crossover line instead of a transition line.

the temperature from the critical temperature and h is the magnetic field. The total free energy F of the system is a function of these parameters:

$$F(t, h, g, \ldots, L).$$

Let us suppose that the system is invariant under a renormalization transformation of scale b and that the parameters transform as

$$t \to tb^{y_t}, \quad h \to hb^{y_h}, \quad g \to gb^{y_g}, \quad \ldots \tag{6.61}$$

While, in general, we must redefine the parameters to make these transformations exact, here we simply assume that (6.61) is satisfied. Then, we say that y_p ($p = t, h, g, \ldots$) is the scaling eigenvalue of the field p and that the scaling dimension of the operator Q conjugate to p is $x_Q \equiv d - y_p$. For example, the asymptotic behavior of the two-point correlation function is

$$\langle Q(x)Q(y) \rangle \propto \frac{1}{|x-y|^{2x_Q}}.$$

If $y_p > 0$, we say that the scaling field p (or the operator Q) is *relevant*. If $y_p < 0$, on the other hand, it is *irrelevant*.

When the parameters obey the scaling transformation (6.61), the free energy satisfies

$$F(t, h, g, \ldots, L) = F(tb^{y_t}, hb^{y_h}, gb^{y_g}, \ldots, L/b).$$

By setting $L/b = \Lambda$ and eliminating b, we get

$$F(t, h, g, \ldots, L) = F(t(L/\Lambda)^{y_t}, h(L/\Lambda)^{y_h}, g(L/\Lambda)^{y_g}, \ldots, \Lambda).$$

If we now regard Λ as a constant, we can eliminate it and obtain the *finite-size scaling* formula (Fisher and Barber, 1972)

$$F(t, h, g, \ldots, L) = \tilde{F}(tL^{y_t}, hL^{y_h}, gL^{y_g}, \ldots). \tag{6.62}$$

From this expression, the scaling form in the thermodynamic limit follows by first deforming (6.62) to

$$f(t, h, g, \ldots, L) = F/L^d = t^{d/y_t} \tilde{f}(tL^{y_t}, ht^{-y_h/y_t}, gt^{-y_g/y_t}, \ldots)$$

and then taking the limit $L \to \infty$ with fixed t to produce the following formula for the free-energy density:

$$f(t, h, g, \cdots) = t^{d/y_t} f(ht^{-y_h/y_t}, gt^{-y_g/y_t}, \ldots). \tag{6.63}$$

The final step is justified because f must converge to some finite value independent of L. Equation (6.63) is called the *scaling form of the free energy*.

By considering a large enough L, we can neglect the dependence on the irrelevant parameters. For example, in the case of the critical point of the two-dimensional Ising model, we have only two relevant scaling fields, one associated with the temperature difference from the critical temperature $T - T_c$ and the other with the magnetic field H. Therefore, we identify t and h in (6.62) with these fields to obtain an asymptotically correct scaling form,

$$F(T, h, L) \approx F(tL^{y_t}, hL^{y_h}) \quad \text{(Ising model)}. \tag{6.64}$$

The scaling forms of other quantities follow from this equation by differentiating both sides with respect to the relevant scaling fields. For example, for the zero-field magnetic susceptibility, differentiating (6.64) twice with respect to H (or h) yields

$$\chi(T, L) \approx -L^{-d} \left. \frac{\partial^2 F((tL^{y_t}, hL^{y_h})}{\partial h^2} \right|_{h=0}$$

$$= L^{-d+2y_h} \tilde{\chi}((T - T_c)L^{y_t}) \quad \text{(Ising model)}, \tag{6.65}$$

where $\tilde{\chi}(X) \equiv -(\partial^2 F(X,Y)/\partial Y^2)_{Y=0}$. From this thermodynamic relation, it follows that $\gamma = (2y_h - d)/y_t$, where γ is the critical exponent characterizing the divergence of the susceptibility at the transition point.

Equations such as (6.64) and (6.65) are of crucial importance in analyzing the results of a numerical simulation. We can use them to identify the fixed point that dominates the phase transition by fitting the computed results to these formulas. For example, suppose we carry out a Monte Carlo simulation of the Ising model for various values of T and L. As a result we will obtain many estimates of $\chi(T, L)$. If we then plot $\chi(T, L)/L^{2y_h-d}$ against $(T - T_c)L^{y_t}$, all data points, regardless of the system size, will fall onto a single curve, $Y = \tilde{\chi}(X)$, provided that L is sufficiently large. Of course, in most cases, we do not know the correct values of y_t, y_h, and T_c. Therefore, we make some guesses. To obtain the correct values, we then adjust the guessed values until all the data points collapse onto a single curve.[20]

With two modifications, we can apply the finite-size scaling discussed above to quantum systems. The first modification is switching from $F \equiv -kT \log Z$ to $\Phi \equiv -\log Z$. This is a natural step, because in the finite-size scaling the imaginary-time axis is treated just as another direction, and therefore we should remove the normalization factor $1/\beta$ to make the equivalence between space and time apparent. The second is to introduce the dynamical critical exponent z, which is the dimension of imaginary time.

In the case of the transverse-field Ising model discussed in the previous section, the imaginary-time axis is equivalent to any of the spatial axes apart from some constant, which means *Lorentz invariance*, $z = 1$. In such cases, we can regard imaginary time as an additional length scale, and when we investigate the quantum critical phenomena, we simply apply the finite-size scaling discussed above with d replaced by $d + 1$. In doing so, we need to fix the aspect ratio by varying the inverse temperature β with the system size L. However, in some cases, the imaginary-time direction is not equivalent to a space direction. To extend the finite-size scaling analysis to such cases, we generalize (6.62) by adding the inverse temperature as another argument,

$$\Phi(t, \ldots, \beta, L) = \Phi(tL^{y_t}, \ldots, \beta/L^z), \tag{6.66}$$

where $t = q - q_c$ measures the distance from the quantum critical point $q = q_c$. This equation describes the zero-temperature phase transition even though it contains a finite inverse temperature. By fitting the simulation data to these forms, we obtain the dynamical scaling exponent z as well as y_t and q_c.

---

[20] Even if we choose the exact values of T_c, y_h, and y_t, the data collapse will not be perfect in practice. This imperfection is due to *corrections to* scaling that arises because we use raw variables such as T and H, not the scaling fields. One can, in principle, address this problem by including correction terms in the fitting.

6.6 Applications to Bosonic systems

In this section, we briefly discuss the application of the algorithms for quantum spin lattices to lattice models of Bosons. Interest in studying Boson lattice models was sparked by various low-temperature experiments (Anderson et al., 1995) on electromagnetically trapped atoms that observed Bose-Einstein condensation. This phenomenon, along with the phenomenon of superfluidity, had long been modeled by systems in a continuum, and, in fact, quantum Monte Carlo simulations in the continuum have been used to study both phenomena for quite some time. The game changer was the experiments realizing an *optical lattice* that traps Bosonic atoms in periodic potential wells created by standing electromagnetic waves (Greiner et al., 2002). The result is an almost ideal realization of the Bose-Hubbard lattice model: While the trapped Bosonic atoms condense, they remain very dilute. The diluteness makes a δ-function description of the inter-atomic interaction plausible, and due to the overlap of the wave functions, the atoms hop (tunnel) from one lattice well to another. (In the optical lattices, the mean distance between the ultra-cold atoms, however, is a thousand times larger than the mean distance between nearest-neighbor lattice sites in a solid.)

The Bose-Hubbard model

$$H = -t \sum_{\langle ij \rangle} (b_i^\dagger b_j + \text{h.c.}) - \frac{U}{2} \sum_i n_i(n_i - 1) - \mu \sum_i n_i, \qquad (6.67)$$

with $n \equiv b_i^\dagger b_i$, is the simplest interacting Boson system relevant to the optical lattice systems. When considered as an approximation of a continuous-space system, it represents a system with short-range repulsive interactions where only the *s*-wave scattering is relevant. However, instead of being an approximation to the continuum, the model is more directly a representative of Bosonic atoms trapped in an optical lattice.

It is straightforward to use the worm algorithm in Section 6.4.2 to study the Bose-Hubbard model with different dimensionalities, lattice geometries, inhomogeneities, and so on. Suppose the local basis is the product basis of local Boson occupation eigenstates $|n_i\rangle$ and that the source operator is defined as $Q_i = b_i^\dagger + b_i$. We then have three states mutually reachable by the worm creation/annihilation process: The state with no worm (0 state) and the two states with the worms. In the worm states, the local occupation number between the head and the tail is larger (+ state) or smaller (− state) than in the original state. According to the general prescription presented in Section 6.4.2, we tune the weight $w(0)$ of the no-worm state for our convenience and make the weights of the worm states

$$w(+) = |\langle n + 1|Q_i|n\rangle| = n + 1 \text{ and } w(-) = |\langle n - 1|Q_i|n\rangle| = n.$$

initial state / weight	final state: n	final state: $n-1$ / n	final state: $n+1$ / n
n — weight 1	0	0	1
$n-1$ / n — weight n	0	0	1
$n+1$ / n — weight $n+1$	$\dfrac{1}{n+1}$	$\dfrac{n}{n+1}$	0

Figure 6.11 Creation and annihilation of a worm in a lattice Boson system in the occupation number representation. Shaded cells correspond to the prohibited transitions. The triangles and the rectangles in the diagrams represent heads and tails, respectively, and each triangle points in the direction of the head's motion (always upward in the examples shown here).

By setting $w(0) = 1$, we easily obtain

$$w(0, +) = w(+, 0) = 1, \quad w(-, +) = w(+, -) = n,$$

as a solution satisfying (6.12). (The other weights vanish.) This solution yields

$$p(+|0) = 1, \quad p(0|+) = \frac{1}{n+1}, \quad p(+|-) = 1, \quad p(-|+) = \frac{n}{n+1},$$

as summarized in Fig. 6.11. In particular, creation and annihilation happen for the $+$ worm (the bottom row or last column of Fig. 6.11) with probability 1 and $1/(n+1)$. The $-$ worms are never created or annihilated by direct processes. They appear or disappear as a result of the head's motion.[21]

The algorithm enables a precise estimation of various quantities of the Bose-Hubbard model as well as other Bosonic models on lattices. For example, we might want to know the most basic property of the model, that is, its phase diagram. To find it amounts to calculating the critical chemical potential μ_c for a given hopping constant t, and we can employ the finite-size scaling method discussed in Section 6.5.3 to numerically locate critical points. To execute the scaling, we carry out Monte Carlo simulations at various values of the control parameter (μ, in the present case). However, when the corresponding term commutes with the

---

[21] Suppose a worm is created as a + worm (e.g., the top-right diagram of Fig. 6.11), and the head moves upward as it first leaves the original position. After spending some time elsewhere, the head can approach the tail from the opposite direction. In such a case, the configuration near the head corresponds to the middle-left diagram of Fig. 6.11 with n replaced by $n + 1$. It thus appears as a $-$ worm.

whole Hamiltonian, as the chemical potential term does in the present case, the task is simplified. In such cases, the control term corresponds to a well-defined quantum number, and we can classify all states according to its value. Therefore, for an arbitrary μ^*, the energy of the lowest state characterized by the quantum number N is

$$E_N(\mu) = E_N(\mu^*) - (\mu - \mu^*)N.$$

If the ground state is in the sector characterized by $N = N_0$ and its excited state is in the $N = N_1$ sector, the excitation gap is

$$\Delta(\mu) = \Delta(\mu^*) - (\mu - \mu^*)\Delta N$$

where $\Delta N \equiv N_1 - N_0$. Suppose μ_c is the critical value and the gap closes for some N_1. Then, by choosing $\mu^* = \mu_c$, we obtain

$$\Delta(\mu) = -(\mu - \mu_c)\Delta N.$$

This equation tells us that to estimate μ_c, we only need to estimate $\Delta(\mu)$ at some μ, namely, $\mu_c \equiv \mu + \Delta(\mu)/\Delta N$. The estimate should be independent of our choice of μ.

To use this physics, we must at least know for which quantum number N_1 the gap closes at the anticipated critical point. In the case of the Mott-insulator/superfluid phase transition in the Bose-Hubbard model, N is the total number of particles, and $N_0 = nV$ where V is the number of sites in the n-th Mott insulator phase. Suppose we increase the chemical potential starting from some point inside the n-th Mott region. As soon as we enter the superfluid phase, the system becomes compressible, and we do not need extra energy to squeeze additional particles into the system. This compressibility also means that as we increase the particle number, the gap eventually vanishes at μ_c, and we can take $N_1 = N_0 + 1$. Therefore, the corresponding gap $\epsilon_p(\mu) \equiv E_{N+1} - E_N = -(\mu - \mu_c)$ is the particle excitation energy at μ, and $\mu_c \equiv \mu + \epsilon_p(\mu)$ is the phase boundary to the superfluid phase in which particles are condensed. Likewise, $\epsilon_h(\mu) \equiv E_{N-1} - E_N = (\mu - \mu_c')$ is the hole excitation energy at μ, and $\mu_c' \equiv \mu - \epsilon_h(\mu)$ is the phase boundary to the superfluid phase in which holes are condensed. The distinction between the particle-condensed states and the hole-condensed states vanishes at large t, and the two states are actually connected to each other without a phase boundary.

Figure 6.12 presents the phase diagram of the Bose-Hubbard model on a simple cubic lattice obtained by a Monte Carlo simulation (Kato and Kawashima, 2010) with an algorithm essentially equivalent to the one discussed previously. The
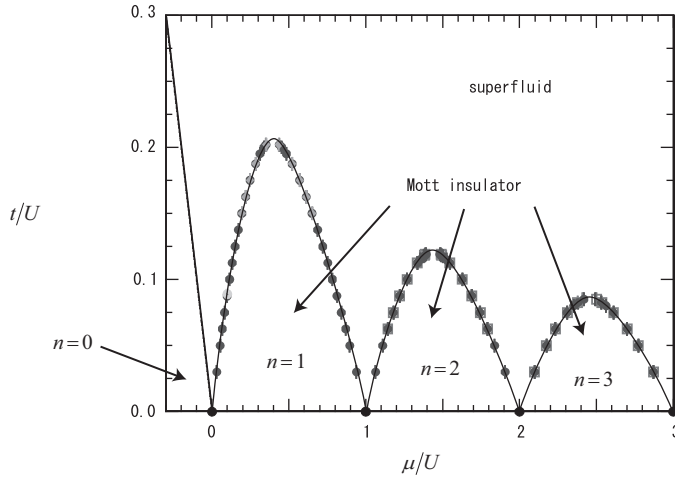
Figure 6.12   The zero-temperature phase diagram of the Bose-Hubbard model on a simple cubic lattice, based on the results of Monte Carlo simulation presented in Kato and Kawashima (2010).

excitation gap was estimated by studying the imaginary-time dependence of the correlation function with

$$\langle b(q,\tau)b^\dagger(q,0)\rangle \propto e^{-\epsilon_p(q,\mu)\tau}$$

for the particle excitation energy at the wave number q, and

$$\langle b^\dagger(q,\tau)b(q,0)\rangle \propto e^{-\epsilon_h(q,\mu)\tau}$$

for the hole excitation energy. For Fig. 6.12, three systems of linear size $L = 8, 12$, and 16 were studied at sufficiently low temperatures so that no significant temperature dependence is visible on the scale adopted in the figure. The size dependence and the statistical error is less than the line thickness.

Similar Monte Carlo techniques are being used in simulations more closely related to real optical-lattice experiments with a trapping potential, that is, a spatially inhomogeneous chemical potential. A recent example is the comparison between experiment and simulation that both trapped approximately 3×10^5 particles (Trotzky et al., 2010). The momentum distribution function was measured and computed, and they agreed with each other at various temperatures and at different potential depths. These results demonstrate the predictive power of the Bose-Hubbard model simulations for optical lattice problems.

There are a number of physical applications of the Bose-Hubbard model or its variants, other than cold atoms in an optical lattice, for which quantum Monte

Carlo simulations are uncovering the rich physics of the model. For example, pair-superfluidity is a phenomenon similar to superconductivity: Two bosonic particles of different type form a quasi-particle, and the quasi-particles condense while single particles do not. In other words, when we have two kinds of particles represented by b_1 and b_2, $\langle b_1 b_2 \rangle \neq 0$, whereas $\langle b_1 \rangle = \langle b_2 \rangle = 0$. Supercounter-flow is a similar phenomenon in which b_2 is replaced by $b_2^\dagger$; namely, particles of one kind are accompanied by holes of the other to form condensing quasi-particles. Bose systems may exhibit the pair-superfluidity or the supercounter-flow state when the repulsive force between the same kind of particles is sufficiently strong (Kuklov and Svistunov, 2003). Simulations (Kuklov et al., 2004) found these phenomena in a two-component Bose-Hubbard model on the square lattice. Supersolidity refers to a state in which a density wave (diagonal order) and a supercurrent (off-diagonal order) coexist. Although such a state has been sought for a long time since the first theoretical speculation (Andreev and Lifshitz, 1969), no conclusive evidence has been obtained so far in any real physical system (Kim and Chan, 2012). However, simulations have found that the hard-core ($U = \infty$) Bose-Hubbard model with nearest neighbor repulsion on the triangular lattice clearly exhibits the coexistence of the two orders (Boninsegni and Prokof'ev, 2005). A simple picture of this phenomenon is that the majority of the particles sustains the crystalline component, while the excess particles act as interstitials that carry the super-current. This simple two-component picture, however, does not seem to hold for the Bose-Hubbard model with nearest-neighbor repulsion on the cubic lattice. A Monte Carlo simulation shows super-solidity even at half-filling (Ohgoe et al., 2012) where no interstitials or excess particles exist except those created by fluctuations.

Suggested reading

M. P. A. Fisher, P. B. Weichman, G. Grinstein, and D. S. Fisher, "Boson localization and the superfluid-insulator transition," *Physical Review B* **40**, 546–70 (1989).

D. M. Ceperley, "Path integrals in the theory of condensed helium," *Review of Modern Physics* **67**, 279 (1995).

J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, 1996).

N. Kawashima and K. Harada, "Recent developments of worldline Monte Carlo methods," *J. Phys. Soc. Japan* **73**, 1379 (2004).

A. Leggett, *Quantum Liquids: Bose Condensation and Cooper Pairing in Condensed-Matter Systems* (Oxford University Press, 2006).

A. W. Sandvik, "Computational studies of quantum spin systems," in *Lectures on the Physics of Strongly Correlated Systems XIV*, ed. A. Avella and F. Mancini, AIP Conference Proceedings, **1297**, 135 (2010).

Exercises

6.1 **Detailed Balance in the General Loop/Cluster Update.** Show an example of a loop/cluster update in which the detailed balance condition does not hold when viewed as a Markov process in the joint space $\Sigma \times \Gamma$.

6.2 **Ising Triangle.** Consider three Ising spins coupled antiferromagnetically to each other and apply the Swendsen-Wang algorithm to it. If the temperature is much lower than the coupling constant, how soon will the system relax to the equilibrium distribution?

6.3 **Loop/Cluster Update in the Ising Limit.** Consider the $S = \frac{1}{2}$ Ising-like *XXZ* model ($J^{xy} \gg J^z > 0$) at zero field. Obtain an explicit formula for the density of graph elements. Show that in the limit $J^{xy}/J^z \to 0$ the algorithm reduces to the conventional Swendsen-Wang algorithm.

6.4 **Griffiths Inequality.** Using Fortuin and Kasteleyn's graphical representation of the partition function of the ferromagnetic Ising model and the graphical observable of the two-point correlation function, prove that the correlation function $\langle S_i S_j \rangle$ always increases when we increase the coupling constant J^z_{kl}, regardless of the choice of i, j, k, and l. Compose a similar proposition for quantum spin models and prove it.

6.5 **Spinons in the One-Dimensional Ising Model.** Compute the expectation value of the energy for the one-dimensional Ising model of length L ($\gg 1$) with periodic boundary conditions. Note that the energy gap is J^z as the first excited state has two domain boundaries and each of them costs $Jz/2$. (Here, we use the $S_i = \pm\frac{1}{2}$ convention.) Confirm (6.59) when $0 < T \ll J^z/\log L$. Also confirm that (6.59) holds for $\Delta E = J^z/2$ instead of J^z for $J^z/\log L \ll T \ll J^z$. How can this apparent paradox be solved?

6.6 **A Single-Site Problem.** Suppose we want to apply the algorithm described by Fig. 6.11 to a single-site problem. Let the effect of the chemical potential be taken into account in the form of the probability for reaching the next scatterer. In the present case, the next scatterer is the tail. Therefore, assume that the probability of reaching the tail, increasing the particle number by one, is $e^{-\beta\mu}$. (The move that decreases the particle number is always accepted.) By direct inspection, show that this procedure produces the correct particle number distribution: $p_n \propto e^{-\mu n}$.