THE PROJECTOR METHOD: THEORY AND EXAMPLES*

EDWARD DENNING DAHL

Stanford Linear Accelerator Center
Stanford University
Stanford, California 94305

January 1985

---

* Ph.D. Thesis.

Table of Contents

List of Illustrations

1. INTRODUCTION

Our work has explored an innovative way of numerically analyzing lattice gauge theories. Previously, most particle theorists have used the Monte Carlo technique introduced by Creutz.[1] However, results obtained using this method are not definitive because of inherent difficulties in the method (stringent computation requirements and relative inefficiency). Our techniques promise potentially more accurate measurements of physically relevant quantities in these same models. Additionally we will clarify the relationship of our technique with the standard one.

We developed the Projector Method technique so it would take advantage of certain simplifying features of gauge theory models. Simultaneous and independent efforts to analyze models in atomic, molecular and condensed matter physics led people to develop the Green Function Monte Carlo.[2] Differences in formalism between our work[3] and the latter line of research have obscured their similarities. We believe these techniques are different applications of one underlying idea. Unifying these approaches may lead to a more complete understanding of the numerical methods and ultimately , the physical models.

The Projector Method possesses several advantages over the usual Monte Carlo algorithm. Since the Projector starts from the Hamiltonian and Hilbert space structure of the problem, the lattice for realistic models is three dimensional. Because of the way that samples are generated in the Projector Method, the critical slowing associated with the standard method is avoided. Finally, the presence of certain conservation laws in the Hamiltonian formulation allow the extraction of string tension between charges that are separated by a distance equal to the size of the lattice.

Starting from a very general notion of what the Projector Method is, we begin applying the techniques to several model problems in chapters 2 and 3. After these examples have traced the development of the actual algorithm from the general principles of the Projector Method, a direct comparison between the

Projector and the Euclidean Monte Carlo is made in chapter 4. Following this is a discussion of the application to Periodic Quantum Electrodynamics in 2 and 3 spatial dimensions. Several methods for improving the efficiency of the Projector in various circumstances are outlined in the final chapter.

Finally, it is hoped that the reader will not be discouraged by the demands made by the author in understanding arguments and technical points discussed in the following pages.

2. APPLICATION TO A MODEL PROBLEM

We present here the most naive type of Projector Method evaluation of an analytic expression. Let M be some $n \times n$ matrix of positive real numbers, and let v be an n dimensional vector of positive real numbers. Suppose we wish to compute $u \equiv M^r v$. Multiplication of M times v takes n^2 multiplications and n^2 additions. Computation of u requires $r \cdot n^2$ multiplications and $r \cdot n^2$ additions. If n and r are small finite numbers, an exact calculation is easy.

If M is the transfer matrix of some statistical mechanics system, then n may be quite large even if the statistical mechanics system is relatively small (in 2 dimensions, the 10×10 site classical Ising model is naturally associated to a 1024 dimensional transfer matrix; in 3 dimensions, the $10 \times 10 \times 10$ site classical Ising model has $n \cong 10^{30}$). Even if we had a computer sufficiently rapid to carry out the multiplication of such a sparse matrix by some vector, most of the $2 \cdot r \cdot n^2$ operations would be trivial since transfer matrices are quite sparse. This inefficient use of computer resources will be avoided if the method of evaluation of the required expression could take advantage of the triviality of most of the arithmetic operations.

A stochastic method may then be appropriate. The elements of the result vector u are given by

$$u_{i_r} = \sum_{i_0, i_1, i_2, \ldots, i_{r-1}} M_{i_r, i_{r-1}} M_{i_{r-1}, i_{r-2}} \cdots M_{i_1, i_0} v_{i_0} \quad . \qquad (2.1)$$

The idea is to sample terms from the sum stochastically. This means that we imagine creating a probability space Ω, each element of which is a sequence of indices. Each element of Ω is associated uniquely with a single term from the sum. Sampling the probability space then allows us to approximately evaluate the sum.

2.1 Choice of measure and random variables

We can dispense entirely with the subtleties of the σ-algebra Σ defined on Ω (see appendix A) and take Σ to be the set of all subsets of Ω, because n is finite. To construct a probability space we now have to define a measure function on Σ. Since each element ω of Ω comprises a set $\{\omega\}$ contained in Σ, we may define the measure function μ on the elements of ω directly and extend it to all of Σ by using the requirement of additivity. The two constraints remaining on the measure μ are that it be non-negative and normalized to 1. It is clear that there are an infinite number of measures we could define to satisfy these constraints, thus we need some criteria to make a reasonable choice. To this end, we jump ahead a bit and explain how to use the measure once it is defined.

Assuming that we have defined some measure on the space Ω, we wish to define a set of random variables U_i ($i = 1, \ldots, n$) so that the expectation of U_i gives u_i. By choosing points from the space Ω according to the probability measure μ, we obtain a sample realization of each of the random variables U_i. From these samples we will compute the means for each U_i to get approximate values of u_i. In order to establish reasonable criteria for choosing a probability measure, note that once the measure is chosen there is only one natural choice for the random variable U_i. This is because we require the expectation of the random variable U_i to be the same as u_i.

To see this, consider the following simple example.

$$Y \equiv \sum_{i \in I} y_i \quad ; \quad y_i \geq 0 \quad . \tag{2.2}$$

We imagine I to be a probability space[*] with measure given by p_i. Thus

---

[*] Here and hereafter we assume the σ-algebra associated to any probability space to consist of the set of all subsets of the probability space if we do not explicitly specify otherwise.

$$p_i \geq 0$$

$$\sum_{i \in I} p_i = 1 \quad .$$
(2.3)

We define a random variable s_i and we want its expectation to be Y.

$$\langle s \rangle = \sum_{i \in I} s_i \, p_i = Y \quad .$$
(2.4)

Given the p_i and Y there are many random variables which satisfy this condition, but one choice is more natural than the others:

$$s_i \equiv \frac{y_i}{p_i} \quad .$$
(2.5)

Note several important features of this example: (i) Choice of the measure dictates a natural choice for the score. (ii) So that s is well-defined, we should never take $p_i = 0$ if $y_i \neq 0$. (iii) If the probability distribution of s is tightly peaked, the method will converge more quickly. This last statement will be a strict requirement on the measure p. This is because choice of measure leads to a natural choice for the score random variable, whose probability function is in turn determined.

Return to our previous example and consider the choice of a measure μ for the probability space Ω of sequences of indices. The standard Projector Method measure we define as follows:

$$P_{ij} \equiv \frac{M_{ij}}{\sum_i M_{ij}} \qquad ; \qquad Q_i = \frac{v_i}{\sum_i v_i}$$
(2.6)

$$\mu \left(i_0, i_1, \ldots, i_r \right) = Q_{i_0} \, P_{i_1, i_0} \, P_{i_2, i_1} \, \cdots \, P_{i_r, i_{r-1}} \quad .$$
(2.7)

The positivity of the measure μ follows from the positivity of the matrix elements of M and of the vector v. It is also easy to see that the normalization condition

5

is satisfied. Consider

$$\int_\Omega d\mu \equiv \sum_{i_0,i_1,\ldots,i_r} \mu\left(i_0,i_1,\ldots,i_r\right) = \sum_{i_0} Q_{i_0} \sum_{i_1} P_{i_1,i_0} \sum_{i_2} P_{i_2,i_1} \cdots \sum_{i_r} P_{i_r,i_{r-1}} \quad .$$

$$(2.8)$$

The sum over i_r may be done using Eq. (2.6). The sum over i_{r-1} is now easily done. This process proceeds until all sums are done and the far r.h.s. of the above expression equals 1.

As indicated above, now that we have chosen a measure for the space Ω of sequences of indices, there is only one natural choice for the random variable whose expectation is a component of the vector u. Before giving this choice, one more technical preliminary must be dealt with. Suppose we are trying to determine a particular component of the resultant vector u, say u_3. If we use the measure μ defined earlier, then many of the elements picked from Ω will not have their final index equal to 3. Such states should not contribute to a computation of the third component of u. Thus the domain of definition of the random variable U_3 whose expectation is u_3 cannot be taken to be all of Ω, but only those elements of Ω for which the final index equals 3. But if the measure μ assigns positive probability to states of Ω outside of the domain of definition of U_3, then a problem arises: the random variable U_3 is not normalized to unity. This means that the sum of the probabilities for all the possible independent outcomes of U_3 is less than 1.

$$\int_{domain\ U_3} d\mu < 1 \quad . \tag{2.9}$$

To avoid this unpleasant state of affairs, we can use a conditional measure for the random variable U_j given by

$$\mu_{i_r=j}\left(i_0,i_1,\ldots,i_{r-1}\right) = \frac{\mu\left(i_0,i_1,\ldots,i_{r-1},j\right)}{\int_{domain\ U_j} d\mu} \quad . \tag{2.10}$$

We are now in a position to define the random variable U_j whose expectation

6

gives the jth component of the resultant vector, u_j.

$$U_j\left(i_0, i_1, \ldots, i_{r-2}, i_{r-1}, j\right) \equiv \frac{M_{j,i_{r-1}}\, M_{i_{r-1},i_{r-2}} \cdots M_{i_1,i_0}\, v_{i_0}}{\mu_{i_r=j}\left(i_0, i_1, \ldots, i_{r-1}, j\right)} \quad . \qquad (2.11)$$

According to our earlier discussion we define U_j only on those sequences that end on the index j. The assertion is that the expectation of U_j is u_j:

$$\int\limits_{\text{domain } U_j} U_j\; d\mu_{i_r=j} = u_j \quad . \qquad (2.12)$$

Use Eqs. (2.10,11) to rewrite the integral in Eq. (2.12).

$$\int\limits_{\text{domain } U_j} U_j\; d\mu = \sum_{i_0,i_1,\ldots,i_{r-1}} \mu_{i_r=j}\left(i_0, i_1, \ldots, i_{r-1}\right) \cdot \frac{M_{j,i_{r-1}} \cdots M_{i_1,i_0}\, v_{i_0}}{\mu_{i_r=j}\left(i_0, i_1, \ldots, i_{r-1}\right)}$$

$$= \sum_{i_0,i_1,i_2,\ldots,i_{r-1}} M_{j,i_{r-1}}\, M_{i_{r-1},i_{r-2}} \cdots M_{i_1,i_0}\, v_{i_0}$$

$$(2.13)$$

The expectation of U_j has the desired form given in Eq. (2.1).

Because of the special form of U, it is simple to construct an algorithm for computing U which only requires knowledge of the components of some sequence ω in a serial fashion. This assertion follows from the definition of U_j in Eq. (2.11) and $\mu_{i_r=j}$ in Eq. (2.10):

$$U_j\left(\omega\right) \equiv \left(\sum_{\nu_0} v_{\nu_0}\right)\left(\sum_{\nu_1} M_{\nu_1,\omega_0}\right)\left(\sum_{\nu_2} M_{\nu_2,\omega_1}\right) \cdots \left(\sum_{\nu_r} M_{\nu_r,\omega_{r-1}}\right) \qquad (2.14)$$

$$j = \omega_r \quad .$$

Knowledge of the first p components of ω allows computation of the leftmost $p+1$ factors in parentheses on the r.h.s. of the equation. Computation of the $p+2^{\text{nd}}$ factor in parenthesis requires only ω_{p+1}, thus all previous components of ω may be discarded. This means that computation of U may be accomplished by keeping track of only the current component of ω.

7

2.2 SAMPLE REALIZATION AND THE SIMULATION FUNCTION

Now that we have defined the probability space Ω and the set of random variables U_j such that the expectations of U_j give u_j, we face the sample realization problem. As discussed in the mathematical appendix, we may phrase the requirement as follows. Given a particular probability space R which we will define in a moment, construct a random variable M on R whose range is not the real numbers, but is the set Ω. The probability space R is the cartesian product of n copies of the unit interval, the σ-algebra on R is just the set of Borel sets on R, and the measure is just the Lebesgue measure in n dimensions, which we denote μ_L. The requirement we place on M is that

$$\mu_L\left(M^{-1}(\sigma)\right) = \mu(\sigma) \tag{2.15}$$

where σ is some element of the σ-algebra Σ over Ω.

The intuitive meaning of this construction is quite simple. We give an example of the construction to clarify the meaning of the requirement above and the simulation function M. In our example Ω is replaced by T, a probability space with a small finite number of elements. Order the elements of T, so they are referred to as t_i, where $i = 1, \ldots, N$. Associated with each t_i is its measure $\mu(t_i)$. Now define a T-valued random variable M on the probability space I, the unit interval with Lebesgue measure.

$$M(x) = t_\ell \quad ; \quad \ell \equiv \sup\left\{ j \,\left|\, \sum_{i<j} \mu(t_i) \,<\, x \right.\right\} \tag{2.16}$$

To understand why we call this a simulation of the probability space T, look at Eq. (2.15) as it applies to this example. Let σ be $\{t_\ell\}$. Then we have

$$M^{-1}(\sigma) = \left(\sum_{i=1}^{\ell-1} \mu(t_i) \,,\, \sum_{i=1}^{\ell} \mu(t_i) \right) \quad . \tag{2.17}$$

Now certainly $\mu_L(a, b) = b - a$ for $b > a$, so

$$\mu_L \left(\sum_{i=1}^{\ell-1} \mu(t_i) \; , \; \sum_{i=1}^{\ell} \mu(t_i) \right) = \sum_{i=1}^{\ell} \mu(t_i) - \sum_{i=1}^{\ell-1} \mu(t_i) = \mu(t_\ell) \quad . \qquad (2.18)$$

Thus the meaning of the simulation function M can be seen as follows: the probability of choosing an element of I in the interval $\left(\sum_{i=1}^{\ell-1} \mu(t_i) \; , \; \sum_{i=1}^{\ell} \mu(t_i) \right)$ is $\mu(t_\ell)$. Whenever we choose an element of I in this interval, and only then, do we say we have chosen the element t_ℓ of T. But the probability of choosing t_ℓ from T is also $\mu(t_\ell)$. Thus M has been constructed so that these two probabilities are equal.

After this brief discussion of the sample realization problem, we may return to the problem of defining an appropriate simulation function M for the space of indices summed over in Eq. (2.1). The first impulse is to use the simulation function M of the most recent example, replacing the space T by Ω. This is generally inadequate for a very simple reason. Computation of the function M from its definition requires a linear search through a significant fraction of the elements of T (or Ω). If T has many elements, this is clearly very inefficient. Even ordering the elements of T so that $\mu(t_i) \geq \mu(t_{i+1})$ is of little help in most cases.

The observation that the sample realization technique employed above uses a linear search order suggests that different types of search orders may be more efficient. A tree-like search through the elements of T is certainly much more efficient: using n yes/no decisions arrayed in a linear fashion we can choose between n objects, using a tree structure we may choose between 2^n objects (see Fig. 1) after making n choices.

In fact the sample realization method of the Projector Method is analagous to the tree-type search, with one important modification. Instead of using the same random number x to determine the decision to be made in each of the boxes, a number x_1 is used to made the decision at the top level, a second random number
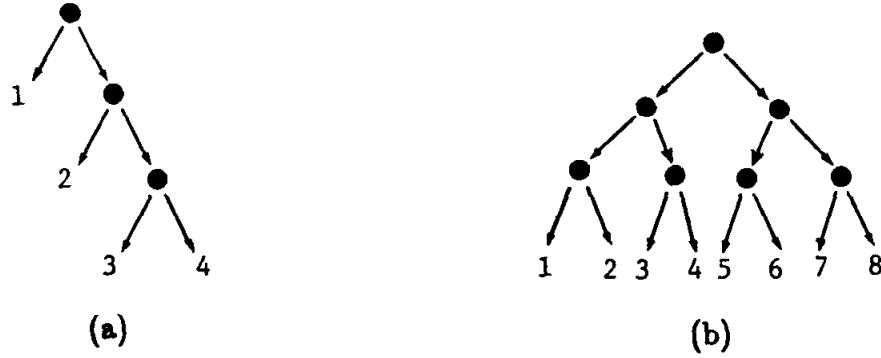
Figure 1. (a) Sequential search order (b) Tree-type search

is used in the appropriate box on the second level, etc. In this way the domain of M is I^n instead of I.

Now define the simulation function M and show that it satisfies Eq. (2.15). The domain of M will be the $r+1$ dimensional unit hypercube, denoted by I^{r+1}. The σ-algebra on I^{r+1} is the set of all Borel sets on I^{r+1} and the measure is Lebesgue measure. Given a point in I^{r+1}, the map M produces some sequence of $r+1$ indices. Denote a generic point in I^{r+1} by X, which has Cartesian coordinates X_i, $i = 0, \ldots, r$. Denote an element of Ω by ω, which is a sequence of indices, the jth of which is ω_j. Now suppose

$$M(X) = \omega \quad . \tag{2.19}$$

Then we have relations between the X_i and the ω_j which define the map M. The set of these relations is

$$\omega_0 = \sup \left\{ j : \sum_{i<j} Q_i < X_0 \right\}$$

$$\tag{2.20}$$

$$\omega_n = \sup \left\{ j : \sum_{i<j} P_{i,\omega_{n-1}} < X_n \right\} , \quad n = 1, \ldots, r .$$

The definitions are inductive, and ω_n depends on X_i for all $i \leq n$.

Now verify Eq. (2.15). It suffices to check the relation with σ replaced by an arbitrary element ω of Ω^*. We check the relation by defining the set R by

$$R \equiv M^{-1}(\omega) \quad . \tag{2.21}$$

Find the Cartesian coordinates of R and then show that its Lebesgue measure equals $\mu(\omega)$.

First verify that the following specification of R agrees with the definition of R given in Eq. (2.21).

$$R \equiv \left\{ X \in I^{r+1} : \sum_{i < \omega_0} Q_i < X_0 < \sum_{i \le \omega_0} Q_i \right. ,$$

$$\left. \sum_{i < \omega_n} P_{\omega_n,\omega_{n-1}} < X_n < \sum_{i \le \omega_n} P_{\omega_n,\omega_{n-1}} \right\} \quad . \tag{2.22}$$

To check this assertion, we will take this last formula as a definition of R and check the following: (i) $X \in R \Rightarrow M(X) = \omega$. (ii) $X \notin R \Rightarrow M(X) \neq \omega$. First check (i). Let $M(X) = \nu$ and ν has components ν_i, $i = 0, 1, \ldots, r$. Show that $\nu_i = \omega_i$, beginning with $i = 0$. According to the definitions in Eqs. (2.20),

$$\nu_0 \equiv \sup \left\{ j : \sum_{i < j} Q_i < X_0 \right\}$$

$$X_0 \in \left(\sum_{i < \omega_0} Q_i , \sum_{i \le \omega_0} Q_i \right) \quad . \tag{2.23}$$

Now from the second line,

$$\sum_{i < \omega_0} Q_i < X_0 \quad , \tag{2.24}$$

therefore ω_0 is an element of the set whose supremum is ν_0. Thus $\nu_0 \ge \omega_0$. Again

---

⋆ The countable additivity of the measure functions allow one to extend the relation from single elements of Ω to subsets of Ω.

from the second line,

$$X_0 < \sum_{i \le \omega_0} Q_i = \sum_{i < \omega_0 + 1} Q_i \le \sum_{i < \omega_0 + m} Q_i \quad , \quad m \ge 1 \tag{2.25}$$

so $\omega_0 + m$ cannot be a member of the set with supremum ν_0. Thus ω_0 is the supremum of the set, and $\nu_0 = \omega_0$.

Now prove the inductive step. Suppose $\omega_i = \nu_i$ for $i = 0, \dots, n-1$. Then we wish to show that $\omega_n = \nu_n$. Again, the relevant definitions are

$$\nu_n \equiv \sup \left\{ j : \sum_{i<j} P_{i,\omega_{n-1}} < X_n \right\}$$

$$X_n \in \left(\sum_{i<\omega_n} P_{\omega_n,\omega_{n-1}} , \sum_{i \le \omega_n} P_{\omega_n,\omega_{n-1}} \right) \quad . \tag{2.26}$$

As before , the relation

$$\sum_{i<\omega_n} P_{\omega_n,\omega_{n-1}} < X_n \tag{2.27}$$

implies that ω_n is in the set whose supremum is ν_n, so that $\nu_n \ge \omega_n$. And again, the relation

$$X_n < \sum_{i \le \omega_n} P_{\omega_n,\omega_{n-1}} = \sum_{i < \omega_n + 1} P_{\omega_n,\omega_{n-1}} \le \sum_{i < \omega_n + m} P_{\omega_n,\omega_{n-1}} \tag{2.28}$$

shows that $\omega_n + m$ cannot be in the relevant set for $m \ge 1$. Thus ω_n is the supremum of the set, and $\omega_n = \nu_n$.

Now we check (ii). If $X \notin R$ then either X_0 lies outside the appropriate interval, or X_n does for some $n = 1, \dots, r$. Suppose X_0 is the offending coordinate. If the lower bound is violated clearly $\nu < \omega_0$ so $\nu \ne \omega$. If the upper bound is violated then $\nu_0 > \omega_0$ and again $\nu \ne \omega$. Analogous reasoning holds if the offending coordinate is not the zeroth. Thus our assertion that the definitions of R in Eqs. (2.21,22) are equivalent is proved.

12

Finally, to check the relation

$$\mu_L\left(M^{-1}(\omega)\right) = \mu(\omega) \tag{2.29}$$

compute the Lebesgue measure of R.

$$\mu_L(R) = \left(\sum_{i \leq \omega_0} Q_i - \sum_{i < \omega_0} Q_i\right) \cdot \prod_{n=1}^{r} \left(\sum_{i \leq \omega_n} P_{\omega_n, \omega_{n-1}} - \sum_{i < \omega_n} P_{\omega_n, \omega_{n-1}}\right) \tag{2.30}$$

$$= Q_{\omega_0} \cdot \prod_{n=1}^{r} P_{\omega_n, \omega_{n-1}} \quad .$$

This is the same as $\mu(\omega)$, originally defined in Eq. (2.7). Thus the simulation function has the desired property expressed in Eq. (2.15).

Through the artifice of the simulation function, the problem of computing expectations of random variables can be looked at somewhat differently. Remember that M is just a function from I^{r+1} to the probability space Ω being simulated, such that

$$\mu_L\left(M^{-1}(\sigma)\right) = \mu(\sigma) \ . \tag{2.31}$$

Now a random variable U defined on Ω is just a measurable real-valued function on Ω. The expectation of U is given by

$$\langle U \rangle = \int_{\Omega} U \ d\mu \quad . \tag{2.32}$$

We can compose M with U to form a random variable $\mathcal{U}$ defined on I^{r+1}:

$$\mathcal{U} \equiv u \circ M \quad . \tag{2.33}$$

By virtue of the relation between M, μ, and μ_L the expectation of $\mathcal{U}$ satisfies

$$\langle \mathcal{U} \rangle_{\mu_L} = \langle U \rangle_{\mu} \quad . \tag{2.34}$$

Thus the expectation value of interest is nothing but the Lebesgue integral over the $r + 1$-dimensional hypercube of the constructed function $\mathcal{U}$.

3. THE ISING QUANTUM SPIN SYSTEM

We apply the formalism developed in chapter 2 to a specific physics problem. The problem is a standard example from quantum statistical mechanics, part of whose usefulness lies in its susceptibility to analytic methods. The model is specified by a Hamiltonian which contains as free parameters the real number h and the integer N.

$$H = -\sum_{\langle ij \rangle}^{N} \sigma_i^{(z)}\sigma_j^{(z)} - h\sum_{i=1}^{N} \sigma_i^{(x)} \quad . \tag{3.1}$$

The Hamiltonian H is an operator in a 2^N dimensional Hilbert space. The relations between the Pauli matrices are standard:

$$\left. \begin{array}{l} \sigma_i^{(x)}\sigma_i^{(x)} = 1 \\ \sigma_i^{(z)}\sigma_i^{(z)} = 1 \\ \sigma_i^{(x)}\sigma_i^{(z)} = -\sigma_i^{(z)}\sigma_i^{(x)} \end{array} \right\} \quad i = 1,\ldots,N \tag{3.2}$$

$$\left. \begin{array}{l} \sigma_i^{(x)}\sigma_j^{(x)} = \sigma_j^{(x)}\sigma_i^{(x)} \\ \sigma_i^{(z)}\sigma_j^{(z)} = \sigma_j^{(z)}\sigma_i^{(z)} \\ \sigma_i^{(x)}\sigma_j^{(z)} = \sigma_j^{(z)}\sigma_i^{(x)} \end{array} \right\} \quad i \neq j, \quad i,j = 1,\ldots,N \quad . \tag{3.3}$$

In extending our treatment of the problem in chapter 2 (where we evaluated $M^r v$) to the Ising model, the vector v will be some state in the Hilbert space. The matrix M will be an approximate version of the exponentiated Hamiltonian. Several issues arise in applying the method to this case. One wishes to multiply the matrix by the vector so that only local computations (ones that do not grow faster than N) have to be made. Also it is preferable not to store the entire probability matrix or score matrix, as the size of these grows exponentially with N. With these guidelines in mind, we begin development of the Projector Method applied to the Ising model.

Pick the matrix M by which we multiply the vector v. If M is the matrix $e^{-\beta H}$ then v is projected onto the ground state of H as β grows larger, provided that the overlap of v with the ground state is non-zero. Since we cannot compute $e^{-\beta H}$, we approximate it with an expression provided by the Trotter product formula:

$$e^{A+B} = \lim_{n \to \infty} \left(e^{A/n} \, e^{B/n} \right)^n \quad . \tag{3.4}$$

Taking n to be some fixed positive integer gives the approximation. A careful choice of A and B makes the problem yield to analysis. For the Ising Model we choose

$$A = -\sum_{\langle ij \rangle}^{N} \sigma_i^{(z)} \sigma_j^{(z)} \quad ; \qquad B = -h \sum_{i=1}^{N} \sigma_i^{(x)} \quad . \tag{3.5}$$

Both A and B are easily exponentiated, so that the r.h.s. of Eq. (3.4) is easily evaluated.

Referring to the initial example in which $M^r v$ is evaluated, we see that a generalization is required to accommodate the Trotter product formula. We now anticipate computing an expression of the form

$$u \equiv MN \ MN \ \ldots \ MN \, v \quad . \tag{3.6}$$

The substance of the work done in chapter 2 is unchanged. We may define a probability space and random variables whose expectations give the components of u. Again the measure is defined as the product of a string of matrices, but now the string alternates between $P(M)$ and $P(N)$, where

$$P(R) \equiv \frac{R_{ij}}{\sum_i R_{ij}} \quad ; \quad R = M, N \quad . \tag{3.7}$$

The issue of defining the simulation function changes only in some detailed particulars. Most importantly, we retain the property that the sequence of basis states ω need not be remembered in its entirety, since the computation of ω and the random variables both can proceed serially.

Using the machinery developed in chapter 2, we now describe a computation of the following quantity carried out in the Ising Model:

$$d \equiv \left\langle \chi \left| \left(e^{-A\beta/n} \, e^{-B\beta/n} \right)^n \right| \phi \right\rangle \quad . \tag{3.8}$$

The ket $|\phi\rangle$ corresponds to the vector v in the earlier example, and the string of operators depends on the positive integer n and the real parameter β. The interpretation of d will depend on n and β. We are not calculating a vector in Hilbert space like in our earlier example. Instead a single matrix element is obtained by taking the overlap of the vector with the bra $\langle \chi |$. We now define and construct the probability space, the random variable, the simulation function, and a simple algorithm for computing the random variable.

In analogy with our earlier example, the structure of the probability space is suggested by writing the number we wish to compute as a sum. By inserting a resolution of the identity operator to the left of each operator $e^{-A\beta/n}$ and $e^{-B\beta/n}$ and to the left of $|\phi\rangle$, we obtain

$$d = \sum_{\omega_{2n+1},\ldots,\omega_0} \langle \chi | \omega_{2n} \rangle \, \langle \omega_{2n} | e^{-\beta A/n} | \omega_{2n-1} \rangle \ldots \langle \omega_2 | e^{-\beta A/n} | \omega_1 \rangle \, \langle \omega_1 | e^{-\beta B/n} | \omega_0 \rangle \, \langle \omega_0 | \phi \rangle \; .$$

$$\tag{3.9}$$

Elements of the probability space will be sequences of Hilbert space basis vectors and each sequence is of length $2n + 1$. Each basis state is stored as a set of N numbers, which are the eigenvalues of the basis state under the N operators $\sigma_n^{(z)}$. Thus an element of the probability space is a sequence of $2n + 1$ sets, each of which is comprised of N numbers, 1 or -1. The generic element of the probability space Ω is ω, the jth element in the sequence ω is ω_j $(j = 0, 1, \ldots, 2n)$, and the eigenvalue of $\sigma_r^{(z)}$ in the state ω_j is $\omega_j(r)$. Note that ω is naturally stored as a two index array of the numbers 1, -1. However, since we have shown that the simulation function and random variables may both be computed serially, we will never have occasion to store the entire array ω at once.

Begin the definition of the measure by choosing the ket $|\phi\rangle$ which will replace v. To motivate our choice for $|\phi\rangle$, observe that the operator product occurring in Eq. (3.8) becomes diagonal as $h \to 0$. In this limit the operator product projects onto two states: one in which the eigenvalue of $\sigma_n^{(z)}$ is 1 for all n, and one in which it is -1. Arbitrarily pick $|\phi\rangle$ to be equal to the first state, so that in the $h \to 0$ limit, the projection process should leave $|\phi\rangle$ invariant. The measure can now be written down in analogy with the previous example.

$$\mu(\omega) = Q_{\omega_0} \, P_{\omega_1,\omega_0}^{(B)} \, P_{\omega_2,\omega_1}^{(A)} \, \ldots \, P_{\omega_{2n},\omega_{2n-1}}^{(A)} \, . \tag{3.10}$$

Q is vector, since it assigns a number (a probability) to each basis vector:

$$Q_{\omega_0} \longleftrightarrow \frac{\langle \omega_0 | \phi \rangle}{\sum_\nu \langle \nu | \phi \rangle} \tag{3.11}$$

$$Q_{\omega_0} \equiv \prod_{r=1}^{N} \delta_{1,\omega_0(r)} \tag{3.12}$$

$$\sum_{\omega_0} Q_{\omega_0} = 1 \; ; \qquad Q_{\omega_0} \geq 0 \, . \tag{3.13}$$

Each of the P's is a matrix whose indices run over the set of basis vectors.

$$P_{\omega_1,\omega_0}^{(A)} \longleftrightarrow \frac{\langle \omega_1 | e^{-\beta A/n} | \omega_0 \rangle}{\sum_\nu \langle \nu | e^{-\beta A/n} | \omega_0 \rangle} \, . \tag{3.14}$$

A is diagonal in this basis, so

$$P_{\omega_1,\omega_0}^{(A)} = \prod_{r=1}^{N} \delta_{\omega_1(r),\omega_0(r)} \, . \tag{3.15}$$

The matrix associated with B is less trivial.

$$P_{\omega_2,\omega_1}^{(B)} \longleftrightarrow \frac{\langle \omega_2 | e^{-\beta B/n} | \omega_1 \rangle}{\sum_\nu \langle \nu | e^{-\beta B/n} | \omega_1 \rangle} \, . \tag{3.16}$$

Using the commutativity of all $\sigma_n^{(z)}$ and the standard identity for exponentiating

Pauli matrices, obtain the following:

$$e^{-\beta B/n} = \prod_{r=1}^{N} \left(1 \cosh\left(\frac{\beta h}{n}\right) + \sigma_r^{(x)} \sinh\left(\frac{\beta h}{n}\right) \right) . \qquad (3.17)$$

The factorizability of this exponential is a key point in the Projector Method, for it essentially allows all computations in the algorithm to be local on the lattice. This simplification motivated our split of the Hamiltonian into two parts.

Understanding how the computation of $P^{(B)}$ is carried out should suffice in convincing the reader that the entire algorithm is now local. The matrix element of $e^{-\beta B/n}$ between $\langle\nu|$ and $|\omega_1\rangle$ can be factored into a product of single site matrix elements:

$$\langle\nu|e^{-\beta B/n}|\omega_1\rangle = \prod_{r=1}^{N} \left\langle \nu(r) \left| \left(1 \cosh\left(\frac{\beta h}{n}\right) + \sigma_r^{(x)} \sinh\left(\frac{\beta h}{n}\right) \right) \right| \omega_1(r) \right\rangle . \quad (3.18)$$

Summing over the 2^N values of $\langle\nu|$ in the denominator of Eq. (3.16) is equivalent to summing over the two values of $\langle\nu(r)|$ for each r:

$$\sum_{\nu} \langle\nu|e^{-\beta B/n}|\omega_1\rangle = \prod_{r=1}^{N}$$

$$\times \left(\sum_{\nu(r)=1,-1} \left\langle \nu(r) \left| \left(1 \cosh\left(\frac{\beta h}{n}\right) + \sigma_r^{(x)} \sinh\left(\frac{\beta h}{n}\right) \right) \right| \omega_1(r) \right\rangle \right) .$$

$$(3.19)$$

The factor in parentheses just gives $\cosh\left(\frac{\beta h}{n}\right) + \sinh\left(\frac{\beta h}{n}\right)$, so we have

$$\sum_{\nu} \langle\nu|e^{-\beta B/n}|\omega_1\rangle = \prod_{r=1}^{N} \left(\cosh\left(\frac{\beta h}{n}\right) + \sinh\left(\frac{\beta h}{n}\right) \right) = e^{Nh/n} . \qquad (3.20)$$

The numerator in Eq. (3.16) is

$$\langle\omega_2|e^{-\beta H/\beta}|\omega_1\rangle = \prod_{r=1}^{N} \left\langle \omega_2(r) \left| \left(1 \cosh\left(\frac{\beta h}{n}\right) + \sigma_r^{(x)} \sinh\left(\frac{\beta h}{n}\right) \right) \right| \omega_1(r) \right\rangle$$

$$= \left[\cosh\left(\frac{\beta h}{n}\right) \right]^j \left[\sinh\left(\frac{\beta h}{n}\right) \right]^{N-j} . \tag{3.21}$$

$$j \equiv \sum_{r=1}^{N} \delta_{\omega_2(r),\omega_1(r)} \tag{3.22}$$

The final expression for $P^{(B)}$ is

$$P^{(B)}_{\omega_2,\omega_1} = p^j \left(1 - p\right)^{N-j} \tag{3.23}$$

$$p = \frac{\cosh(\frac{\beta h}{n})}{\exp(\frac{\beta h}{n})} \quad . \tag{3.24}$$

$P^{(B)}$ has the form of a probability for N independent distinct events, each of which has one outcome of probability p and one of probability $(1 - p)$. The locality of the algorithm manifests itself in the fact that transitions made at each lattice site are independent of one another.

Positivity of the measure follows from positivity of the individual matrix elements. The verification that

$$\sum_{\omega} \mu(\omega) = 1 \tag{3.25}$$

is carried out exactly as in the discussion after Eq. (2.8) and follows from

$$\sum_{\omega_{r+1}} P^{(A)}_{\omega_{r+1},\omega_r} = 1 \; ; \qquad \forall \; \omega_r \tag{3.26}$$

$$\sum_{\omega_{r+1}} P^{(B)}_{\omega_{r+1},\omega_r} = 1 \; . \tag{3.27}$$

Now define a random variable D whose expectation gives d. In analogy with

Eq. (2.5),

$$D(\omega) \equiv \langle x|\omega_{2n}\rangle\langle\omega_{2n}|e^{-\beta A/n}|\omega_{2n-1}\rangle\langle\omega_{2n-1}|e^{-\beta B/n}|\omega_{2n-2}\rangle \ldots \langle\omega_2|e^{-\beta A/n}|\omega_1\rangle$$

$$\times \; \langle\omega_1|e^{-\beta B/n}|\omega_0\rangle\langle\omega_0|\phi\rangle \Big/ \mu(\omega) \; .$$

(3.28)

Note that unlike U_j in Eq. (2.11), D is defined on all of Ω. The computation of the expectation of D gives the matrix element defined in Eq. (3.8).

$$\langle D\rangle_\mu \equiv \int D(\omega) \; d\mu(\omega) = \sum_\omega \mu(\omega) \cdot D(\omega) = d$$

(3.29)

The definitions of D and d involve the state $\langle\chi|$. Define $\langle\chi|$ so that its overlap with every basis state is unity.

$$\langle x|\nu\rangle = 1 \qquad \forall \; |\nu\rangle \; .$$

(3.30)

This choice is made so that every ω selected from Ω will contribute non-trivially to the expectation of D.

Given the definition of the measure (Eq. (3.10)) and the matrices and vector occurring in the definition of the measure (Eqs. (3.12,15,23)), the definition of the simulation function M is straightforward. There is one subtlety which must be addressed and is connected with the domain of definition of M. In the example of chapter 2, the domain of M was I^{r+1} and not I^1. The analog in the Ising model would be to take the domain of M to be I^{2n+1}. This approach means each of the $2n + 1$ coordinates of a point in I^{2n+1} would govern the transition from a basis state ω_r at the rth step to a basis state ω_{r+1} at the $r + 1^{\text{th}}$ step. Since there are 2^N possibilities for the basis state ω_{r+1}, one random number is used to choose between 2^N possibilities. To avoid the inefficiency of using a single random number to select between many alternatives, we adopt a tree-like search pattern for the new basis state that uses N random numbers. We enlarge the domain of M from I^{2n+1} to $I^{N(2n+1)}$. Each coordinate of a point in the domain of the simulation function is required to make a binary decision.

We now begin the explicit definition of M. Let X be a point in $I^{(2n+1)N}$ and denote its Cartesian coordinates by X_i, $i = 1, \ldots, N(2n+1)$. Put $M(X) = \omega$ where ω_r is the basis state which is the rth element of the sequence ω. Define M by giving $\omega_r(n)$ in terms of X_i. Because of the form of Q_{ω_0}, the state ω_0 is fixed and the dependence of ω_0 on X_i is trivial:

$$\omega_0(r) = 1 \ . \tag{3.31}$$

For the next two elements of the sequence , we have

$$\omega_1(r) = \omega_0(r) \left[\theta(p - X_{N+r}) - \theta(X_{N+r} - p) \right] \tag{3.32}$$

$$\omega_2(r) = \omega_1(r) \tag{3.33}$$

where p was defined in Eq. (3.24). The coordinates X_1 through X_N are reserved for determining ω_0, thus the coordinates X_{N+1} through X_{2N} are used to get ω_1. Since H_A is a diagonal operator, ω_2 is determined from ω_1 without dependence on the coordinates X_{2N+1} through X_{3N}. Extend the definition of M to specify the remaining components of ω in terms of the coordinates of X.

$$\omega_{2n+1}(r) = \omega_{2n}(r) \left[\theta(p - X_{N(2n+1)+r}) - \theta(X_{N(2n+1)+r} - p) \right] \tag{3.34}$$

$$\omega_{2n}(r) = \omega_{2n-1}(r) \ . \tag{3.35}$$

This completes the definition of the simulation function.

Now check that M satisfies the criteria $\mu_L(M^{-1}(\omega)) = \mu(\omega)$. First compute the l.h.s. as was done for the similar expression in chapter 2. Then simplify the definition of the r.h.s. provided in Eq. (3.10) to verify the desired equality.

If ω does not satisfy Eq. (3.35) for all n, r then ω is not in the range of M and it follows that $M^{-1}(\omega) = \phi$ and $\mu(M^{-1}(\omega)) = 0$. Alternatively, suppose ω

does satisfy Eq. (3.35). Then

$$M^{-1}(\omega) = \{X \in I^{2nN} : \quad X_{2nN+r} \in (0,1) \quad \text{and}$$

$$X_{N(2n+1)+r} \in \left(p \cdot \delta_{\omega_{2n+1}(r),-\omega_{2n}(r)} , \, p + (1-p) \cdot \delta_{\omega_{2n+1}(r),-\omega_{2n}(r)}\right)\} \quad .$$

$$(3.36)$$

This follows directly from the definitions in Eq. (3.34,35). In this case the set $M^{-1}(\omega)$ is a hyper-rectangle and its Lebesgue measure is just the product of its extent in each dimension.

$$\mu_L(M^{-1}(\omega)) = \prod_{t=1}^{(2n+1)N} (U_t - L_t)$$

$$U_t \equiv \begin{cases} 1 & \text{if} \quad t = 2nN + r \\ p + (1-p) \cdot \delta_{\omega_{2n+1}(r),-\omega_{2n}(r)} & \text{otherwise} \end{cases} \qquad (3.37)$$

$$L_t \equiv \begin{cases} 0 & \text{if} \quad t = 2nN + r \\ p \cdot \delta_{\omega_{2n+1}(r),-\omega_{2n}(r)} & \text{otherwise} \end{cases}$$

To simplify this expression , note that

$$U_t - L_t = \begin{cases} 1 & \text{if} \quad t = 2nN + r \\ p^{\alpha_{n,r}} \cdot (1-p)^{\alpha_{n,r}} & \text{otherwise} \end{cases} \qquad (3.38)$$

$$\alpha_{n,r} \equiv \delta_{\omega_{2n+1}(r),-\omega_{2n}(r)}$$

and so

$$\mu_L(M^{-1}(\omega)) = p^{\alpha} \cdot (1-p)^{2nN-\alpha}$$

$$\alpha \equiv \sum_{n,r} \alpha_{n,r} \; . \qquad (3.39)$$

The definition of the r.h.s. in Eq. (3.10) vanishes if ω does not satisfy Eq. (3.35) for all n, r, since the matrices $P^{(A)}$ are diagonal (they are the identity

matrix). Thus we have verified the desired equality in this case. Now suppose ω does satisfy Eq. (3.35) for all n, r. Then

$$
\mu(\omega) = \prod_{l=1}^{2n} \exp\left(\frac{-\beta Nh}{n}\right) \cdot \left[\cosh\left(\frac{\beta h}{n}\right)\right]^{j_l} \cdot \left[\sinh\left(\frac{\beta h}{n}\right)\right]^{N-j_l}
$$

$$
\dot{j}_n \equiv \delta_{\omega_{2n+1}(r), -\omega_{2n}(r)} \cdot
$$

(3.40)

Carry out the product to obtain

$$
\mu(\omega) = \exp\left(-2\beta Nh\right) \cdot \left[\cosh\left(\frac{\beta h}{n}\right)\right]^{J} \cdot \left[\sinh\left(\frac{\beta h}{n}\right)\right]^{2nN-J}
$$

$$
J \equiv \sum_{l=1}^{2n} \dot{j}_l \cdot
$$

(3.41)

Using the definition of p in Eq. (3.24), the quantities in Eq.(3.39) and Eq.(3.41) are identical, and the equality is verified.

Now that the probability space has been defined along with the relevant random variable, and the simulation function has been constructed, we have, in principle, finished describing the computation of the number d. There are several algorithms which implement the computation, and they differ greatly in their efficiency and computation requirements. We outline the naive implementation, then indicate how the computation can be simplified. The simplification results from a property of the random variable (it can be computed serially) and a property of the measure (it represents a Markov process).

The naive implementation of the method begins by picking a point X at random from $I^{N(2n+1)}$. This entails picking the $N(2n+1)$ Cartesian coordinates X_i of X. Using these coordinates compute the image of X under M, which we denote ω. Now from the sequence ω compute and record $D(\omega)$, which is one instance of the random variable D. Repeat this chain of events many times and construct an approximate probability distribution for D from the samples. The mean of the distribution gives an estimate for d, and its scatter about the mean can be used to estimate the uncertainty.

As mentioned above, the Projector Method measure has been chosen to have the form of a Markov process. A Markov process refers to a probability distribution for a sequence of events in which the conditional probability for a particular event (conditioned by specifying the outcome of several earlier events) in the sequence depends only on the most recent event specified. For example specification of any state in the sequence allows determination of the conditional probability distribution for the following state in the sequence. To see that this assertion is true, form the conditional probability distribution for the $j + 1^{st}$ state in the sequence, given that the previous j states are fixed. Take j even for convenience.

$$
\begin{aligned}
P\left(\omega_{j+1}|\omega_0,\ldots,\omega_j\right) &= \frac{\displaystyle\sum_{\omega_{j+2},\ldots,\omega_{2n}} Q_{\omega_0} P^{(B)}_{\omega_0,\omega_1} \cdots P^{(A)}_{\omega_{2n},\omega_{2n-1}}}{\displaystyle\sum_{\omega_{j+1},\ldots,\omega_{2n}} Q_{\omega_0} P^{(B)}_{\omega_0,\omega_1} \cdots P^{(A)}_{\omega_{2n},\omega_{2n-1}}} \\[2em]
&= \frac{\displaystyle\sum_{\omega_{j+2},\ldots,\omega_{2n}} P^{(B)}_{\omega_j,\omega_{j+1}} P^{(A)}_{\omega_{j+1},\omega_{j+2}} \cdots P^{(A)}_{\omega_{2n},\omega_{2n-1}}}{\displaystyle\sum_{\omega_{j+1},\ldots,\omega_{2n}} P^{(B)}_{\omega_j,\omega_{j+1}} P^{(A)}_{\omega_{j+1},\omega_{j+2}} \cdots P^{(A)}_{\omega_{2n},\omega_{2n-1}}} \\[2em]
&= P^{(B)}_{\omega_j,\omega_{j+1}} \, .
\end{aligned}
\tag{3.42}
$$

Thus we see that this conditional probability distribution reduces to the conditional probability distribution that results when only the immediately preceding state in the sequence is fixed.

The related property of the random variable D is serial computability. To see that D may be computed serially, refer to the definitions of D and μ to see that

$$
\begin{aligned}
D(\omega) &= \frac{\langle \omega_{2n}|e^{-\beta A/n}|\omega_{2n-1}\rangle}{P^{(A)}_{\omega_{2n},\omega_{2n-1}}} \cdot \frac{\langle \omega_{2n-1}|e^{-\beta B/n}|\omega_{2n-2}\rangle}{P^{(B)}_{\omega_{2n-1},\omega_{2n-2}}} \cdots \frac{\langle \omega_2|e^{-\beta A/n}|\omega_1\rangle}{P^{(A)}_{\omega_2,\omega_1}} \\[2em]
&\quad \cdot \frac{\langle \omega_1|e^{-\beta B/n}|\omega_0\rangle}{P^{(B)}_{\omega_1,\omega_0}} \cdot \frac{\langle \omega_0|\phi\rangle}{Q_{\omega_0}} \, .
\end{aligned}
\tag{3.43}
$$

The argument that demonstrated the property of serial computability in the

random variable U_j (see the discussion concerning Eq. (2.14)) applies identically to the above expression.

The fact that the measure on Ω describes a Markov process means that elements may be chosen from Ω according to their probability by selecting their components serially. Thus the choice of ω_j depends only on ω_{j-1}. This is reflected in the recursive structure of the simulation function (see Eqs. (3.32-5)). The fact that D is serially computable means that an algorithm exists for its computation which is compatible with the serial selection of components of Ω.

The simplified algorithm picks[*] the first N Cartesian coordinates of X to select ω_0. The right-most factor of $D(\omega)$ is then calculated. Now pick the next N coordinates of the point X. Using the definition of ω_1 in terms of ω_0 and the new coordinates we determine ω_1. Next calculate the factor second to the right in the definition for D. These two right-most factors may be multiplied, since the functional dependence of D on these arguments is so simple. Consideration of the logical extension of this procedure shows that the state ω_0 will not be needed any further. Thus, we may overwrite ω_0 in storage and decrease the storage requirements on the computer. This process is extended until all the components of the sequence ω have been computed and the random variable has been evaluated at this point.

The interplay between the properties of the measure and random variables should be clear by now. It is this structure which allows the explicit simulation of the probability space that motivated the development of the Projector method. It is an added bonus that this method allows the reduction in storage requirements from 4 to 3 dimensions for realistic models.

---

[*] The probability distribution for the point X factors into independent probability distributions for each of its Cartesian coordinates.

4. COMPARISON: PROJECTOR
AND STANDARD MONTE CARLO

We elucidate some general principles of the Projector Method which distinguish it from the standard Monte Carlo. We review the basic formalism underlying the standard Monte Carlo. Then the Projector Method is discussed with reference to this formalism.

In the standard Monte Carlo simulation,[4] there is an underlying probability space whose elements correspond to particular configurations of a classical mechanics system. There is a Boltzmann measure defined on the probability space and certain random variables are defined on this probability space. The expectations of these variables give macroscopically measurable quantities such as thermodynamic potentials and specific heats. The Monte Carlo simulation requires that we generate some approximation to the true measure and then obtain approximate expectations of the random variables in the Boltzmann measure by using the expectations in the approximate measure.

Here is how the approximate measure is generated. Suppose μ_B is the exact measure. The standard Monte Carlo provides an algorithm for producing N elements of the probability space for any N such that the following requirement is fulfilled . If state ω is chosen n_ω times, so that

$$\sum_\omega n_\omega = N \ , \tag{4.1}$$

then

$$\lim_{N \to \infty} \frac{1}{N} n_\omega = \mu_B(\omega) \ . \tag{4.2}$$

The above limit must be interpreted in the appropriate probabilistic sense. The choice of the N states yields an approximate measure μ_N which ultimately

converges to μ_B.

$$\mu_N(\omega) \equiv \frac{n_\omega}{N}$$

$$\lim_{N \to \infty} \mu_N = \mu_B \tag{4.3}$$

For example, suppose we wish to calculate the average energy. Then

$$\langle \epsilon \rangle_{\mu_B} \equiv \sum_\omega \epsilon_\omega \, e^{-\beta \epsilon_\omega} \Big/ \sum_\omega e^{-\beta \epsilon_\omega} \; . \tag{4.4}$$

The Boltzmann measure is given by

$$\mu_B(\omega) = \frac{e^{-\beta \epsilon_\omega}}{\sum_\omega e^{-\beta \epsilon_\omega}} \; . \tag{4.5}$$

The standard Monte Carlo algorithm picks N states, with the ωth state occurring n_ω times. The expectation of the random variable ϵ in the approximate measure μ_N is simply

$$\langle \epsilon \rangle_{\mu_N} = \sum_\omega \epsilon_\omega \, \mu_N(\omega) \; . \tag{4.6}$$

If N is very large then

$$\mu_N(\omega) = \mu_B(\omega) \tag{4.7}$$

so that

$$\langle \epsilon \rangle_{\mu_N} = \langle \epsilon \rangle_{\mu_B} \; . \tag{4.8}$$

At this level of abstraction there is no difference between the standard Monte Carlo and the Projector Method. Both rely on simulation of a given probability space and random variables defined on it. The differences first manifest themselves when we analyze the technique used for simulating the probability space.

The Boltzmann measure approximated in standard Monte Carlo suffers from the disadvantage that it cannot be directly calculated. The normalization sum in the denominator is the partition function, which is notoriously intractable. The standard Monte Carlo sidesteps this difficulty by relying on a Markov process to generate the N states which determine the approximate measure. The Markov process generates a sequence of configurations stochastically, so that the probability distribution of one state in the sequence depends on all previous states in a simple way: the dependence is only on the immediately preceding state.

The Projector Method differs radically from the standard Monte Carlo at this point. The measure for the Projector Method probability space is explicitly calculable from Eq. (2.7). The simulation function M can always be calculated explicitly. The Markov process upon which the standard Monte Carlo relies is unnecessary in the Projector Method. A very different Markov process is used in the Projector Method.

We would like to explicitly distinguish between the way in which the standard Monte Carlo and the Projector Method rely on Markov processes. This analysis begins by recognizing that both simulation problems may be divided into two parts. The first part is the sampling problem discussed in the first mathematical appendix: both techniques must somehow "sample the probability space" according to their respective measure. The second part of the problem is to evaluate the random variable at specific elements of the probability space. It is useful to divide each technique into these two parts because our analysis relates only to the first part of the techniques.

Our analysis will consist of transforming a standard Monte Carlo simulation into a Projector Method simulation. Imagine a finite lattice statistical mechanics problem in D dimensions. Single out one dimension as the time direction with coordinate τ and construct a transfer matrix for this direction. Denote the configuration of the $D - 1$ dimensional sublattice specified by $\tau = n$ by ω_n $(n = 0, \dots, N)$. The boundary conditions for the lattice in the $D - 1$

28

directions orthogonal to τ may be chosen in any way. Fix the configuration at $\tau = 0$ and do not allow the introduction of any coupling between the lattice states at $\tau = 0$ and $\tau = N$.

If the above conditions are satisfied, then we write the partition function as

$$Z = \sum_{\omega_1,\ldots,\omega_N} \exp\left\{-\beta\, S(\omega_0, \omega_1, \ldots \omega_N)\right\} \qquad (4.9)$$

where S is the energy or action, depending on one's prejudices. If S does not contain couplings of range longer than nearest neighbor in the τ direction, introduction of the transfer matrix T allows us to rewrite the Boltzmann factor.

$$\exp\left\{-\beta\, S(\omega_0, \omega_1, \ldots \omega_N)\right\} = T_{\omega_0,\omega_1}\, T_{\omega_1,\omega_2}\, \cdots\, T_{\omega_{N-1},\omega_N} . \qquad (4.10)$$

The partition function is related to a product of transfer matrices:

$$Z = \sum_{\omega_1,\ldots,\omega_N} T_{\omega_0,\omega_1}\, T_{\omega_1,\omega_2}\, \cdots\, T_{\omega_{N-1},\omega_N} . \qquad (4.11)$$

The Boltzmann measure, which just gives the probability of obtaining a given state from the canonical ensemble, is given by

$$\mu_B(\omega_0,\ldots,\omega_N) = \frac{T_{\omega_0,\omega_1}\, T_{\omega_1,\omega_2}\, \cdots\, T_{\omega_{N-1},\omega_N}}{Z} . \qquad (4.12)$$

If the sublattice configurations ω_0 through ω_j inclusive are fixed, the probability of sublattice $j+1$ assuming a value ω_{j+1} is given by

$$P\left(\omega_{j+1}|\omega_0,\ldots,\omega_j\right) = \frac{\displaystyle\sum_{\omega_{j+2},\ldots,\omega_N} T_{\omega_0,\omega_1}\, \cdots\, T_{\omega_{N-1},\omega_N}}{\displaystyle\sum_{\omega_{j+1},\ldots,\omega_N} T_{\omega_0,\omega_1}\, \cdots\, T_{\omega_{N-1},\omega_N}}$$

$$= \frac{\displaystyle\sum_{\omega_{j+2},\ldots,\omega_N} T_{\omega_j,\omega_{j+1}}\, T_{\omega_{j+1},\omega_{j+2}}\, \cdots\, T_{\omega_{N-1},\omega_N}}{\displaystyle\sum_{\omega_{j+1},\ldots,\omega_N} T_{\omega_j,\omega_{j+1}}\, \cdots\, T_{\omega_{N-1},\omega_N}} \qquad (4.13)$$

$$\equiv R_{\omega_j,\omega_{j+1}}(j) .$$

It is simple algebra to verify the following identity:

$$\mu_B(\omega_0, \ldots, \omega_N) = R_{\omega_0, \omega_1}(0) \; R_{\omega_1, \omega_2}(1) \; \ldots \; R_{\omega_{N-1}, \omega_N}(N-1) \; . \tag{4.14}$$

Note that the R matrices satisfy

$$R_{\omega_j, \omega_{j+1}}(j) \; > \; 0 \tag{4.15}$$

and

$$\sum_{\omega_{j+1}} R_{\omega_j, \omega_{j+1}}(j) = 1 \; . \tag{4.16}$$

Therefore we may use these R matrices to implement a Projector Method simulation, identically to the way we use the P matrices from Eq. (2.7). The fact that the Boltzmann measure is expressed as a product of R matrices means that a Projector Method simulation using $R(j)$ in place of P will select states according to the Boltzmann measure. The non-trivial dependence of the R matrices on j means that the Markov process employed by the Projector method is non-stationary.

We translate a Projector Method simulation into a standard Monte Carlo simulation to complete our analysis. Begin with a finite spatial lattice and arbitrary boundary conditions. Associated with the lattice is a Hamiltonian operator and a Hilbert space in which the Hamiltonian acts. Specify a basis in the Hilbert space by simultaneously diagonalizing a maximal mutually commuting collection of operators (typically one or several operators per site). A basis state is specified by the eigenvalues of these operators. Elements of the Projector Method probability space are N-element sequences of basis vectors. If we denote the jth basis vector in such a sequence by ω_j, the Projector Method measure is given by

$$\mu_{PM}(\omega_0, \ldots, \omega_N) = Q_{\omega_0} \; P_{\omega_1, \omega_0} \; P_{\omega_2, \omega_1} \; \ldots \; P_{\omega_N, \omega_{N-1}} \tag{4.17}$$

where the construction of P in terms of the Hamiltonian and the basis has been discussed. To translate the simulation procedure of the Projector Method into

30

a standard Monte Carlo simulation, construct a Markov Process which will converge to the above measure. We will not carry out this construction explicitly, but instead will indicate why this is a straightforward exercise.

In order that a Markov Process generates a sequence of states that will converge to some measure, one requires three conditions of the Markov process: normalization, ergodicity, and homogeneity. Normalization requires that the transition from any state be made to some other state with probability unity. Ergodicity means that a finite number of applications of the stochastic matrix suffices to connect any two states. Homogeneity says that the desired measure is an eigenvector of the stochastic matrix with eigenvalue unity.

Given some measure to be simulated, the above requirements do not select a stochastic matrix . Many such matrices obey the above three properties. Two familiar examples are the heat bath and the Metropolis algorithm. In fact any stochastic matrix satisfying the property of detailed balance will converge to the desired measure. An appropriate stochastic matrix may be constructed as follows. Visit each site of the lattice in any predetermined but random order making random changes to the eigenvalues of the operators there. These changes are made with probabilities specified by the principle of detailed balance. These probabilities are not difficult to calculate, since probabilities for each of the possible final states are given explicitly by the measure.

To summarize our comparison of the standard Monte Carlo and Projector Method, we recapitulate. One can simulate a Boltzmann measure for a classical statistical mechanics system in D dimensions using Projector Method ideas. This means choosing elements from the ensemble of systems one at a time by filling out $D - 1$ dimensional sublattices one after another, until the D dimensional lattice is filled out. Then evaluate the random variables, discard the configuration, and begin again. The drawback is the difficulty of calculating the matrices R which specify the probabilities of choosing the various new $D-1$ dimensional sublattices. Going in the other direction is trivial: a Projector Method measure can always

be simulated using a Markov process. An initial sequence of N basis states is chosen in an arbitrary way. Then make systematic and stochastic changes to this sequence of basis vectors according to some stochastic matrix. Naturally the Markov matrix has to be constructed to converge to the Projector Method measure, but we have noted this is possible in practice as well as in principle.

As an application of the above analysis, we simulate the classical two dimensional Ising model by a Projector Method technique. On each site of a two dimensional lattice (see Fig. 2) there is a spin variable which can assume two values: 1 and -1. The Hamiltonian is

$$H = -\sum_{\langle ij \rangle} S_{(i)} S_{(j)} \ . \tag{4.18}$$

The sum is over nearest neighbors and determines the magnetic character of the model.

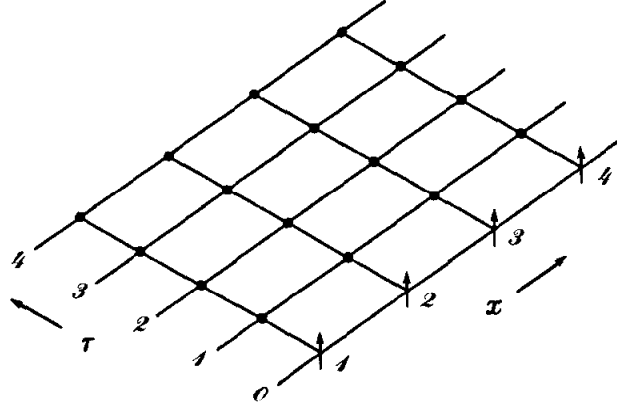**Figure 2.** Ising model lattice

The spin configuration at $\tau = 0$ is fixed. The boundary conditions are periodic in the x direction, and the spins at $\tau = 0$ and 4 have three, not four, nearest neighbors. The measure of a particular spin configuration ω is given by

$$\mu(\omega) \equiv \frac{e^{-\beta H(\omega)}}{Z}$$

$$Z \equiv \sum_{\omega} e^{-\beta H(\omega)} \ . \tag{4.19}$$

To proceed with the Projector Method type simulation we first compute the R matrices of Eq. (4.13). Since configurations will be selected at $\tau = 1, 2, 3, 4$ we compute $R(0)$, $R(1)$, $R(2)$, $R(3)$.

The following recursive definition scheme expresses $R(j)$ in terms of $R(j+1)$ and an additional quantity Z. First define Z:

$$Z_{i_{j-1}}(j-1) \equiv \sum_{i_j} T_{i_{j-1}, i_j} \; Z_{i_j}(j)$$

$$Z_{i_4}(4) \equiv 1 \; . \tag{4.20}$$

The Z functions can be computed recursively in reverse order starting at $j = 3$ and working back to $j = 0$. The Z's are the denominators needed in Eq. (4.13).

$$Z_{i_j}(j) = \sum_{i_{j+1}, \ldots, i_N} T_{i_j, i_{j+1}} \; T_{i_{j+1}, i_{j+2}} \; T_{i_{N-1}, i_N} \tag{4.21}$$

The R matrices are given by

$$R_{i_j, i_{j+1}}(j) = \frac{T_{i_j, i_{j+1}} \; Z_{i_{j+1}}(j+1)}{Z_{i_j}(j)} \tag{4.22}$$

The matrix $R_{i_j, i_{j+1}}(j)$ gives the probability that the state at time slice j will be i_{j+1}, given that the state at time slice $j-1$ is i_j. Using these R matrices, one can pick states from the canonical ensemble according to the Boltzmann measure.

How can the Projector Method measure be improved so that statistical fluctuations are reduced? The inspiration for the following suggestion comes from the last example. There, the Boltzmann measure is factored in a way which allows application of the Projector Method. This technique suggests we can choose new Projector Method measures, as long as they factorize in a similar fashion.

Correspondingly, the new Projector Method measure $\bar{\mu}_{PM}$, is defined as in Eq. (2.7).

$$\bar{\mu}_{PM}(\omega) = P_{\omega_1, I}(1) \; P_{\omega_2, \omega_1}(2) \; \ldots \; P_{\omega_N, \omega_{N-1}}(N) \; . \tag{4.23}$$

Note that the Markov process governed by these P matrices is not stationary,

since the P matrices depend on the current time slice. We have avoided the issue that H must be divided into H_1 and H_2. The basis state I is the initial element of all sequences ω. The P matrices will be defined by a recursive scheme identical to Eqs. (4.20-22). The exponentiated Hamiltonian replaces the transfer matrix.

$$U_{\omega_j, \omega_{j+1}} \equiv \langle \omega_{j+1} | e^{-\beta H / s} | \omega_j \rangle$$

$$Q_{\omega_{j-1}}(j-1) \equiv \sum_{\omega_j} U_{\omega_{j-1}, \omega_j} \, Q_{\omega_j}(j) \tag{4.24}$$

$$P_{\omega_j, \omega_{j+1}}(j) = \frac{U_{\omega_j, \omega_{j+1}} \, Q_{\omega_{j+1}}(j+1)}{Q_{\omega_j}(j)}$$

Random variables can be defined whose expectations in $\bar{\mu}_{PM}$ equal the expectations of D, N, etc. in μ_{PM}. The statistical fluctuations encountered in simulating the first set of random variables should be significantly smaller than those encountered in the second set. This is because I believe the Boltzmann measure more accurately reflects the important terms in the sums that yield the random variables of interest (see Eq.(3.9)). Modifying the Projector Method measure to reflect the properties of a Boltzmann type measure should then mean the important terms in the sum are sampled more frequently. This suggests a direction in which future Projector Method research might profitably develop.

5. PERIODIC QUANTUM ELECTRODYNAMICS

We now duplicate the analysis of the Ising Model and apply it to Periodic Quantum Electro-dynamics (PQED). The Hilbert Space and Hamiltonian of the problem are defined after some geometrical preliminaries are reviewed. The symmetries of the Hamiltonian are mentioned and a basis in Hilbert space is chosen. The probability space and measure are defined as well as the typical random variable. The simulation function is then discussed.

5.1   THE HAMILTONIAN, ITS STATES AND SYMMETRIES

We will apply our formalism to PQED in two and three spatial dimensions, so the notation for vertices, links, and plaquettes of the lattice will be kept general. These parts of the lattice will be referred to as v_i, ℓ_j, and p_k, where $i = 1, \ldots, V$, $j = 1, \ldots, L$, and $k = 1, \ldots, P$. For a two dimensional (three dimensional) lattice with cubic connectivity and toroidal boundary conditions, V, L, and P are respectively $N, 2N, N$, $(N, 3N, 3N)$. Each link is associated with the two vertices which are its endpoints, and each plaquette is associated with the four links which are its edges. The fact that each plaquette has four edges merely reflects the "triangulation" of the spatial torus. The association of vertices to links and links to plaquettes may be made precise[5] through the introduction of a boundary operator ∂ which acts on formal sums of vertices, links, plaquettes, and higher dimensional "simplices", if they exist.

This formalization is not without benefits, for it allows a convenient expression of the lattice versions of geometric differential operators such as curl, grad, div. This is useful to the extent that continuum QED has a simple geometrical interpretation. We will not undertake the automation of formalism necessitated by a careful definition of the boundary operator and concomitant definitions of coincidence numbers and orientation. Instead, we list the formulas we will use and note that a rigorous meaning may be given to the symbols. We will content ourselves with a heuristic explanation.

The boundaries of the vertices, links, and plaquettes are given by

$$\partial v_i = 0$$

$$\partial \ell_j = \sum_i d_1 (j, i) \cdot v_i \qquad (5.1)$$

$$\partial p_k = \sum_j d_2 (k, j) \cdot \ell_j \ .$$

The first equation merely asserts the fact that a point has null boundary. The second equation indicates that the boundary of the jth link is a formal sum over all vertices, weighted by the factor $d_1 (j, i)$. The weight d_1 assumes values $0, 1, -1$ according to the following rules. If vertex v_i is not at one of the two ends of link ℓ_j, then $d_1 = 0$. Now imagine assigning an arrow to each link. If v_i is the vertex pointed at by the arrow on link ℓ_j, then $d_1 = 1$, and if v_i is at the other end of link ℓ_j then $d_1 = -1$.
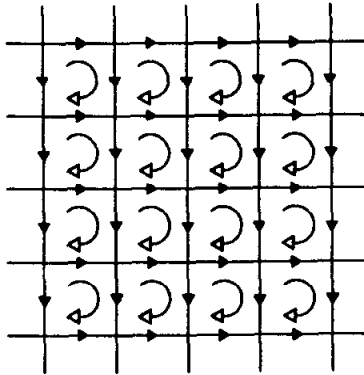
Figure 3. Spatial lattice for 2+1-dimensional PQED

The third equation expresses the boundary of the kth plaquette as a weighted formal sum over links. Weighting factor d_2 assumes the values $0, 1, -1$ and vanishes if ℓ_j is not one of the links bordering the plaquette p_k. To determine d_2 when it is non-vanishing, one must assign an orientation, clockwise or counterclockwise, to each plaquette. If the arrow on a link bordering a plaquette matches

36

the plaquette's orientation, then $d_2 = 1$. Otherwise $d_2 = -1$. The typical orientation assigned to links and plaquettes in the two dimensional case is shown in Fig. 3.

The Hilbert space of the problem can now be defined. It is the space of complex-valued square integrable functions of L variables, where each variable θ runs from $-\pi$ to π. Each link is associated to one of these variables, thus we refer to the variables as θ_{ℓ_j}. We define a pair of operators for each link: θ_{ℓ_j}, $\frac{1}{i}\frac{\partial}{\partial\theta_{\ell_j}}$. The first operator is bounded and may be defined on all of the Hilbert space:

$$\sup_{|\psi\rangle} \frac{\|\theta_{\ell_j} |\psi\rangle\|}{\||\psi\rangle\|} = \pi \ . \tag{5.2}$$

The second operator is unbounded and cannot be defined on all of Hilbert space. We can choose its domain to be the subspace of finite linear combinations of vectors like

$$\prod_{j=1}^{L} \exp\left(i n_{\ell_j} \theta_{\ell_j}\right) \tag{5.3}$$

where n_{ℓ_j} assigns an integer to each link. The operator $\frac{1}{i}\frac{\partial}{\partial\theta_{\ell_j}}$ has finite norm for any vector in this subspace, and this subspace spans the Hilbert space. Self-adjointness of both sets operators on their domains can be verified.

Lastly we wish to check the canonical commutation relations:

$$\left[\frac{1}{i} \frac{\partial}{\partial\theta_{\ell_j}} \, , \ \theta_{\ell_{j'}}\right] = \frac{1}{i} \, \delta_{j,j'} \ . \tag{5.4}$$

Unfortunately the domains of definition of the operators $\frac{1}{i}\frac{\partial}{\partial\theta_{\ell_j}}\theta_{\ell_{j'}}$ and $\theta_{\ell_{j'}}\frac{1}{i}\frac{\partial}{\partial\theta_{\ell_j}}$ do not overlap, so there is no significance to the above relation. The solution is to relinquish the validity of the desired relations, and instead rely on the Weyl relation:

$$\left[\frac{1}{i} \frac{\partial}{\partial\theta_{\ell_j}} \, , \ e^{i\theta_{\ell_{j'}}}\right] = e^{i\theta_{\ell_{j'}}} \, \delta_{j,j'} \ . \tag{5.5}$$

These relations are formally obtained from the canonical commutation relations and possess the advantageous feature of being true on a dense domain of the

Hilbert space. The significance of choosing this particular domain for the derivative operators will become clear when we examine the symmetries of the problem.

The Hamiltonian is a Hermitian operator in the Hilbert space:

$$H = g^2 \sum_{j=1}^{L} \left(\frac{1}{i} \frac{\partial}{\partial \theta_{\ell_j}} \right)^2 - \frac{1}{g^2} \sum_{k=1}^{P} \cos \theta_{p_k}$$

$$\theta_{p_k} \equiv \sum_{j=1}^{L} \theta_{\ell_j} \cdot d_2(k,j) \ .$$

(5.6)

H is unbounded, but it is defined on the same domain on which the derivative operators are defined. The first term is identified with the $\vec{E}^2$ term in the classical Hamiltonian. A power series expansion of the cosine term and retention of the first non-trivial part gives the $\vec{B}^2$ term in the classical Hamiltonian.

The form of H has been largely dictated by a desire to transcribe to the lattice a version of continuum QED which retains explicit gauge invariance. The price paid for this desire is that interactions have been introduced via the cosine term. Truncation of the cosine term at its quadratic part would yield a solvable theory that lacks explicit gauge invariance. Current folklore maintains that the continuum limit of the theory is more likely to be gauge invariant if the regularized form of the theory maintains the gauge invariance. By continuum limit we mean the normal heuristic procedure.[6] A physical length is assigned to the lattice spacing, and this length is taken to 0 while the coupling constant g (and possibly an overall scale factor in the definition of H) is varied so as to maintain constant the correlation length, as measured in physical units. The theory is expected to exhibit Lorentz invariance in the limit that the physical length scale of the lattice vanishes.

Gauge symmetries are manifested in different ways in the final theory, depending on the quantization scheme employed. Here quantization has been carried out in the axial gauge, eliminating the time-like component of the vector

potential as a degree of freedom from the Lagrangian. The passage to the Hamiltonian is then canonical. Examination of the resulting Hamiltonian reveals a large number of symmetries: one per spatial point. These symmetries correspond exactly to the unfixed time-independent gauge degrees of freedom. Diagonalizing the symmetry operators which commute with the Hamiltonian fixes the remaining gauge degrees of freedom, and is equivalent to choosing a configuration of external charges in which the system is to evolve. In our problem these symmetry operators are labelled by sites:

$$G_{v_i} \equiv \sum_{j=1}^{L} -d_1\,(j,i) \cdot \frac{1}{i}\, \frac{\partial}{\partial \theta_{\ell_j}} \ . \tag{5.7}$$

The operator G_{v_i} is just the lattice version of the divergence of the electric field at the vertex v_i. It is straightforward to check that

$$[H\ ,\ G_{v_i}] = 0 \ . \tag{5.8}$$

The natural interpretation of the eigenvalue of G_{v_i} is as the external charge at the site v_i.

The reason for the particular choice of the domain for the electric field operators can now be stated. The spectrum of each G_{v_i} contains 0 in this domain, thus there is a vanishing external charge sector of the Hilbert space. If we had chosen a different domain for the derivative operators, then no longer would there exist a sector of the Hilbert space in which all G_{v_i} would have 0 eigenvalue.

To make this explicit, suppose for the link ℓ_1 we choose the subspace

$$e^{i\alpha\theta} \sum_{n=-\infty}^{\infty} a_n\, e^{i\theta_{\ell_1} n} \ ; \qquad a_n \neq 0 \text{ for finitely many } n \tag{5.9}$$

to be the domain of $\frac{1}{i}\frac{\partial}{\partial \theta_{\ell_1}}$. The spectrum of this operator is shifted from $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$ to $-3+\alpha,\ -2+\alpha,\ -1+\alpha,\ \alpha,\ 1+\alpha,\ 2+\alpha,\ 3+\alpha, \ldots$.

If v_1 (v_2) is oriented positively (negatively) with respect to ℓ_1, then the spectrum of G_{v_1} (G_{v_2}) is now $\ldots, -3 + \alpha, -2 + \alpha, \ldots$ ($-3 - \alpha, -2 - \alpha, \ldots$). Thus the effect of this change of domain is to effectively place a fractional charge $+\alpha$ at v_1 and $-\alpha$ at v_2. We are more interested in the analysis of our numerical method than in analysis of these charges, so this topic will not be pursued further here. These alternative domains are exploited in exhibiting a confinement mechanism by Drell et al (see reference 7).

To complete the analysis of the symmetry properties of our Hamiltonian, note that

$$\sum_{i=1}^{V} G_{v_i} = 0 , \tag{5.10}$$

implying that of the N symmetry operations, only $N - 1$ are independent. The above equation has a natural interpretation, which is that the total external charge on the lattice must vanish: this is obtained by integrating Gauss's Law over the volume of the lattice.

In D spatial dimensions there are D additional symmetry operators which have not been included in the above discussion. These correspond to the possibility of introducing an external background electric field in any of the spatial directions. These symmetry operators can be written as linear combinations of the electric field operators. In two dimensions they are

$$G_x = \sum_{\ell_j \in S_x} \frac{1}{i} \frac{\partial}{\partial \theta_{\ell_j}} \tag{5.11}$$

$$G_y = \sum_{\ell_j \in S_y} \frac{1}{i} \frac{\partial}{\partial \theta_{\ell_j}} . \tag{5.12}$$

See Fig. 4 for the sets S_x and S_y. One can then check that

$$[H , G_x] = [H , G_y] = 0 . \tag{5.13}$$

There is a great deal of arbitrariness in choosing the links to include in S_x or S_y. For instance, choosing the set S'_x of Fig. 4 gives rise to a different operator G'_x.

However, it can be shown that G'_x may be obtained by adding combinations of the G_{v_i} to G_x. The same arbitrariness exists in choosing S_y.
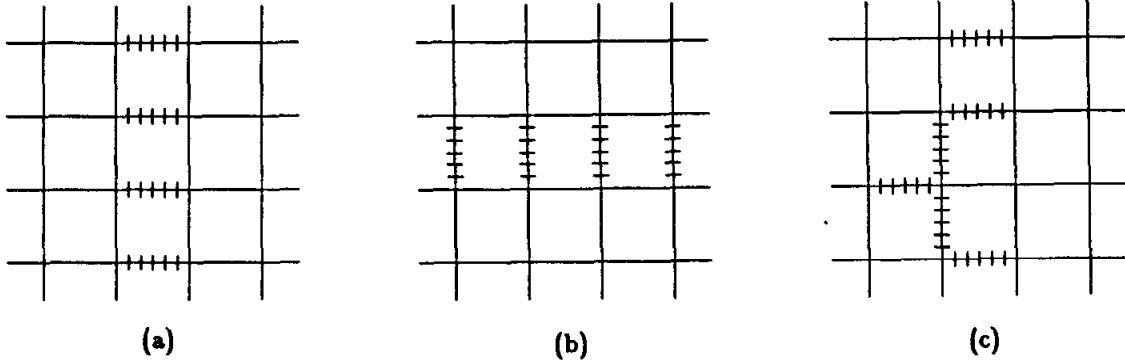
Figure 4. Sets of links used to define background electric field: (a) S_x (b) S_y (c) S'_x

We wish to choose a basis that diagonalizes the Hamiltonian in the strong coupling limit. The basis states are then

$$|n\rangle = (2\pi)^{-L/2} \prod_{j=1}^{L} \exp\left(in_{\ell_j}\theta_{\ell_j}\right) \ . \tag{5.14}$$

It follows that

$$\frac{1}{i} \frac{\partial}{\partial\theta_{\ell_j}} |n\rangle = n_{\ell_j} |n\rangle \ . \tag{5.15}$$

The eigenvalues labeling the basis states are just the electric field values on the links. The vector potential, which is conjugate to the electric field in this formulation, has maximal uncertainty in this basis. Local gauge symmetries formulated above are especially transparent in this basis. We find that the sum of eigenvalues on the inward pointing links at a vertex minus the sum of the eigenvalues on the outward pointing links gives the external charge on a vertex. This quantity vanishes at each vertex in the vanishing external charge sector.

5.2 Choice of Probability Space and Random Variables

We begin the definition of the probability space Ω. As before, its elements are sequences ω of Hilbert space basis vectors. An element of the sequence is ω_n, and the eigenvalue of ω_n under the electric field operator on link ℓ_j is $\omega_n(\ell_j)$. Now we turn to defining the measure on Ω. The form of the measure is identical to that in the Ising model case[*]

$$\mu(\omega) = Q_{\omega_0}\, P^{(B)}_{\omega_1,\omega_0}\, P^{(A)}_{\omega_2,\omega_1} \, \cdots \, P^{(A)}_{\omega_{2n},\omega_{2n-1}} \tag{5.16}$$

though here the P matrices are harder to compute. The matrix elements of P are expressed in terms of matrix elements of the exponentiated Hamiltonian, and thus we embark upon an analysis of this operator.

In two spatial dimensions the Hamiltonian is split into two parts so that each part may be exponentiated exactly. The following decomposition is convenient in that H_A and H_B are treated identically. The electric term is divided evenly between H_A and H_B while the magnetic term is split on a plaquette by plaquette basis into two parts.

$$H_A = \frac{g^2}{2} \sum_{j=1}^{L} \left(\frac{1}{i}\,\frac{\partial}{\partial \theta_{\ell_j}}\right)^2 - \frac{1}{g^2} \sum_{p_k \in P_A} \cos\theta_{p_k} \tag{5.17}$$

$$H_B = \frac{g^2}{2} \sum_{j=1}^{L} \left(\frac{1}{i}\,\frac{\partial}{\partial \theta_{\ell_j}}\right)^2 - \frac{1}{g^2} \sum_{p_k \in P_B} \cos\theta_{p_k} \, . \tag{5.18}$$

This splitting of all plaquettes into the sets P_A and P_B is checkerboard fashion as in Fig. 5.

---

[*] This is true for $2+1$-dimensional PQED. For the $3+1$-dimensional case it is necessary to subdivide the Hamiltonian into four pieces so that each piece may be exponentiated locally.

Figure 5. Decomposition of plaquettes into sets P_A and P_B

The Hamiltonian is a sum of these two parts,

$$H = H_A + H_B , \qquad (5.19)$$

and H_A and H_B are separately gauge invariant. Furthermore, we can write

$$H_X = \sum_{p_k \in P_X} h_{p_k} ; \qquad X = A, B \qquad (5.20)$$

where h_{p_k} is a single plaquette Hamiltonian given by

$$h_{p_k} = \frac{g^2}{2} \sum_{j=1}^{L} \left(d_2\,(k,j) \cdot \frac{1}{i} \frac{\partial}{\partial \theta_{\ell_j}} \right)^2 - \frac{1}{g^2} \cos \left(\sum_{j=1}^{L} d_2\,(k,j) \cdot \theta_{\ell_j} \right) . \qquad (5.21)$$

Note that

$$\left[h_{p_k} , h_{p_{k'}} \right] = 0 \quad \text{if} \quad k, k' \in A \quad \text{or} \quad k, k' \in B . \qquad (5.22)$$

From this follows the fact that H_A and H_B can be exponentiated, provided we can exponentiate the single plaquette Hamiltonian.

In appendix B the single plaquette Hamiltonian h is numerically exponentiated. The Trotter product formula is used to provide a convergent approximation to e^{-h}. Truncated matrices represent the operators in the approximation, and

43

the matrices are handled on a computer. The error in this computation can be made as small as desired, and in our case is comparable to the accuracy in representing a real number on the computer.

The operators e^{-H_A} and e^{-H_B} are tensor products of the single plaquette operators e^{-h}, thus matrix elements of the former pair are appropriate products of matrix elements of the latter. The P matrices appearing in the definition of the measure are defined as in the Ising model example.

$$P^{(A)}_{\omega_{2n},\omega_{2n-1}} = \frac{\langle \omega_{2n} | e^{-\beta H_A/n} | \omega_{2n-1} \rangle}{\sum_\nu \langle \nu | e^{-\beta H_A/n} | \omega_{2n-1} \rangle} \tag{5.23}$$

$$P^{(B)}_{\omega_{2n+1},\omega_{2n}} = \frac{\langle \omega_{2n+1} | e^{-\beta H_B/n} | \omega_{2n} \rangle}{\sum_\nu \langle \nu | e^{-\beta H_B/n} | \omega_{2n} \rangle} \tag{5.24}$$

The sums in the denominators factor into separate sums for each plaquette, and thus can be carried out locally.

As opposed to the Ising model case, neither H_A nor H_B is diagonal in this basis, which has been chosen to diagonalize the strong-coupling limit of the problem. In fact, H_A and H_B are just translations of one another. More precisely, the unitary operators which implement spatial translations can be used to classify operators on the Hilbert space. Under this classification, H_A and H_B form a two-dimensional irreducible representation of the group of translations.

The factor Q which occurs in Eq. (5.16) is a Kronecker delta function of its index ω_0. This has the effect of assigning zero probability to every sequence of states that does not start with the particular state $|\phi\rangle$ specified by the Kronecker delta. The initial state $|\phi\rangle$ may lie in any symmetry sector. Because the operators H_A and H_B commute with the operators which label the external charge and background field, the entire sequence ω will lie in the same sector as ω_0. This leads to the technique of choosing $|\phi\rangle$ to lie in a particular symmetry sector in order to determine information about that sector.

The simplest random variable to simulate is analogous to D, defined in Eq. (3.28) in the Ising model. We defer discussing more sophisticated random variables until several improvements to the basic method have been developed. The expectation of the simplest random variable satisfies

$$\langle D \rangle_\mu = \langle \chi | \left(e^{-\beta H_A/n} e^{-\beta H_B/n} \right)^n | \phi \rangle . \tag{5.25}$$

This quantity decays exponentially in β at a rate equal to the lowest eigenvalue of H (in the sector specified by $|\phi\rangle$). Extracting this rate is the most naive way of determining ground state energies in the Projector Method.

The simulation function M is similar to the one defined for the Ising model. The domain of M is I^{nP}. As noted earlier, neither e^{-H_A} nor e^{-H_B} is diagonal and thus M depends non-trivially on all its coordinates. In the Ising model case, each coordinate of the sample point X was used to make a binary decision. For this case each coordinate corresponds to visiting one plaquette in the lattice. The coordinate is used to pick an integer Δn which determines the change to the plaquette quantum numbers via Eq. (B16). Since the probability for making a transition involving an addition or subtraction of arbitrarily many units of flux is non-zero, the single coordinate is used to make an infinitary decision. Fortunately, for typical values of the couplings, several of the infinite number of possibilities comprise 99% of the range of the coordinate (see Fig. 6). Thus, picking Δn given the Cartesian coordinate is not computationally difficult.
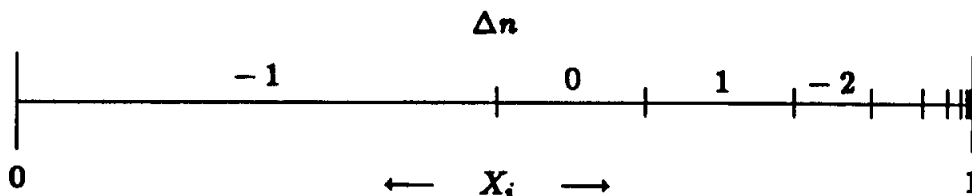
Figure 6. Δn as a function of X_i

Finally, the similarity of this problem to the Ising model problem should convince the reader that the random variable D possesses the property of serial computability. Furthermore, the measure constitutes a Markov process. These properties allow a reduction of the storage requirements to spatial degrees of freedom only.

All these arguments apply equally to the $3 + 1$-dimensional case, except that H must be divided into four parts. It is impossible to reduce the number of sub-Hamiltonians below four since each link borders four plaquettes. The single plaquette Hamiltonians associated with these four plaquettes do not commute with one another. Thus, they must be included in separate sub-Hamiltonians, of which there must then be at least four. That it is actually possible to divide all the plaquette terms into four groups so that the sub-Hamiltonians consist of sums of commuting terms is best seen by construction. We will not carry out this construction but note that it is straightforward.

6. PARALLEL SCORING AND EXTENSIONS

By now the mathematical formalism underlying the Projector Method has been constructed. We will no longer be so explicit in the mathematical exposition since we wish to change the focus of our analysis slightly. Here we present several applications and modifications of the basic method. These include trial parameter, parallel scoring, and multiple configuration techniques. Numerical results are presented.

6.1 PARALLEL SCORE EVALUATION OF ENERGIES

There are several methods for computing ground state energies using the Projector Method. All methods follow more or less directly from the observation that

$$e^{-\delta\beta\epsilon_0} \cong \frac{\langle\chi|e^{-(\beta+\delta\beta)H}|\phi\rangle}{\langle\chi|e^{-\beta H}|\phi\rangle} \,. \tag{6.1}$$

Here $|\phi\rangle$ and $|\chi\rangle$ should have non-zero overlap with the ground state, and β should be large. The simplest method is the $(NE + NT)/NT$ method. A modification of this method yields the parallel score technique, and a particular limit of the parallel score technique gives the derivative technique.

In the $(NE + NT)/NT$ method we define a pair of random variables, the expectations of which give the numerator and denominator of Eq. (6.1). The most naive implementation requires that the numerator be simulated, then the denominator, and then the ratio is formed. The ratio of the expectations of the random variables is the exponential of the mass gap multiplied by $-\delta\beta$.

There exists a simple variant of the $(NT + NE)/NT$ method, for which the method is named. The variant arises from the observation that the work done in computing the denominator of the ratio in Eq. (6.1) can be used in the computation of the numerator. To see this remember that we do not actually simulate matrix elements of $e^{-\beta H}$, but an approximation to this operator. The Trotter product approximation necessitates choice of an integer, in the infinite

47

6. PARALLEL SCORING AND EXTENSIONS

By now the mathematical formalism underlying the Projector Method has been constructed. We will no longer be so explicit in the mathematical exposition since we wish to change the focus of our analysis slightly. Here we present several applications and modifications of the basic method. These include trial parameter, parallel scoring, and multiple configuration techniques. Numerical results are presented.

6.1 PARALLEL SCORE EVALUATION OF ENERGIES

There are several methods for computing ground state energies using the Projector Method. All methods follow more or less directly from the observation that

$$e^{-\delta\beta\epsilon_0} \cong \frac{\langle\chi|e^{-(\beta+\delta\beta)H}|\phi\rangle}{\langle\chi|e^{-\beta H}|\phi\rangle} \,. \tag{6.1}$$

Here $|\phi\rangle$ and $|\chi\rangle$ should have non-zero overlap with the ground state, and β should be large. The simplest method is the $(NE + NT)/NT$ method. A modification of this method yields the parallel score technique, and a particular limit of the parallel score technique gives the derivative technique.

In the $(NE + NT)/NT$ method we define a pair of random variables, the expectations of which give the numerator and denominator of Eq. (6.1). The most naive implementation requires that the numerator be simulated, then the denominator, and then the ratio is formed. The ratio of the expectations of the random variables is the exponential of the mass gap multiplied by $-\delta\beta$.

There exists a simple variant of the $(NT + NE)/NT$ method, for which the method is named. The variant arises from the observation that the work done in computing the denominator of the ratio in Eq. (6.1) can be used in the computation of the numerator. To see this remember that we do not actually simulate matrix elements of $e^{-\beta H}$, but an approximation to this operator. The Trotter product approximation necessitates choice of an integer, in the infinite

limit of which the approximation becomes exact. If we choose the integer to be NT, then every sequence of NT basis states selected using the simulation function gives rise to a particular value of the score random variable relevant to the denominator. Choose $\delta\beta$ so that

$$\frac{\delta\beta + \beta}{\beta} = \frac{NT + NE}{NT} \tag{6.2}$$

for some positive integer NE. Then instead of choosing a new sequence of $NT + NE$ basis states for the numerator, just choose NE additional ones and append them to the initial sequence of NT states. The numerator random variable is evaluated on the full sequence of $NT + NE$ states as though the sequence had been picked in the standard way. We assert that the expectation of the numerator random variable computed in this way is identical to the expectation computed in the normal manner. The proof of this assertion follows directly from Eq. (6.1) and the definition of the simulation functions and scores.

In fact, we can imagine that the entire problem consisting of a pair of probability spaces and their associated formalism collapses into one probability space: the one for the numerator. The numerator random variable is unchanged, but we have to give a definition of the denominator random variable. The new denominator variable evaluated at a point of the numerator probability space is just the original denominator variable evaluated on the first NT elements of the sequence which is the element of the numerator probability space. The new denominator random variable has a probability distribution identical with the original denominator random variable. The high degree of correlation observed between the numerator and denominator random variables suggests that computing these expectations in parallel is a statistical improvement.

We now turn to the parallel score technique. Imagine that we have defined a probability space, a random variable D, and a simulation function. Then the identity

$$\langle D \rangle_\mu = \left\langle \chi \left| \left(e^{-\beta A/s}\, e^{-\beta B/s} \right)^s \right| \phi \right\rangle \tag{6.3}$$

48

is expected to hold. Now define a second random variable N on the probability space with the property that

$$\langle N \rangle_\mu = \left\langle \chi \left| \left(e^{-(\beta+\delta\beta)A/s} \, e^{-(\beta+\delta\beta)B/s} \right)^s \right| \phi \right\rangle . \tag{6.4}$$

The value of N on a particular element ω of the probability space is

$$N(\omega) = \langle \chi | \omega_{2n} \rangle \langle \omega_{2n} | e^{-(\beta+\delta\beta)A/s} | \omega_{2n-1} \rangle \langle \omega_{2n-1} | e^{-(\beta+\delta\beta)B/s} | \omega_{2n-2} \rangle$$
$$\dots \langle \omega_2 | e^{-(\beta+\delta\beta)A/s} | \omega_1 \rangle \, \langle \omega_1 | e^{-(\beta+\delta\beta)B/s} | \omega_0 \rangle \dots \langle \omega_0 | \phi \rangle \Big/ \mu(\omega) \tag{6.5}$$

in analogy with the definition of D. Clearly, N shares the virtue of serial computability with D. The similarity of form between D and N means that the simulation of N can be carried out with very little additional computing overhead if D is already being simulated. The matrix elements used in the definition of N are tabulated identically to the matrix elements used in the definition of D. Thus, all indexing arithmetic done to compute D may be used in computing N.

As with the $(NE + NT)/NT$ method, there is a strong correlation between the random variables D and N. If $\delta\beta = 0$ then the random variables become equal. The ground state energy is computed according to

$$\epsilon_0 = \frac{1}{\delta\beta} \, \ell n \left(\frac{\langle N \rangle_\mu}{\langle D \rangle_\mu} \right) . \tag{6.6}$$

The temptation to make $\delta\beta$ very small is curbed by the presence of the factor $1/\delta\beta$ in the energy. This is called parallel scoring because a single sequence of basis states is converted into a pair of scores which track one another quite closely.

Note that the introduction of the energy scale $\delta\beta$ into the problem seems unnecessary. Elimination of this scale by passing to the limit $\delta\beta \to 0$ yields the derivative method. In order to isolate the random variable upon which this method relies, begin by considering the random variable N of the parallel score

method. Clearly N is parametrized by $\delta\beta$, so we can Taylor expand the random variable N about $\delta\beta = 0$.

$$N(\delta\beta) = N(0) + \delta\beta \cdot N'(0) + O(\delta\beta^2) . \tag{6.7}$$

The above equation is understood as an equality between functions defined on the probability space. Using the fact that $N(0) = D$, we have

$$\begin{aligned}
\epsilon_0 &= \frac{1}{\delta\beta} \, \ell n \left(\frac{\langle D + \delta\beta \, N'(0) + O(\delta\beta^2) \rangle_\mu}{\langle D \rangle_\mu} \right) \\
&\cong \frac{\langle N'(0) \rangle_\mu}{\langle D \rangle_\mu} + O(\delta\beta) .
\end{aligned} \tag{6.8}$$

The energy scale $\delta\beta$ does not appear in the first term of the power series expansion that contains the ground state energy. The price paid for this simplification is the necessity of simulating a new type of variable, $N'(0)$.

To see how the simulation of $N'(0)$ must proceed, recall some basic principles. The underlying probability space has been chosen and the expectation value of $N'(0)$ can be written down as a sum of terms. To define a random variable whose expectation is the desired sum, impose a condition similar to Eq. (2.5).

Let us carry out this prescription out explicitly. The expectation of $N'(0)$ satisfies

$$\langle N'(0) \rangle_\mu = \frac{d}{d\delta\beta} \langle N(\delta\beta) \rangle , \tag{6.9}$$

where we have

$$\langle N(\delta\beta) \rangle = \sum_{\omega_0} \cdots \sum_{\omega_{2n}} \langle \chi | \omega_{2n} \rangle \langle \omega_{2n} | e^{-(\beta+\delta\beta)A/s} | \omega_{2n-1} \rangle \cdots \langle \omega_1 | e^{-(\beta+\delta\beta)B/s} | \omega_0 \rangle \langle \omega_0 | \phi \rangle .$$

$$\tag{6.10}$$

Forming the $\delta\beta$ derivative of this expression and evaluating it at $\delta\beta = 0$ will give the expectation value of $N'(0)$. Doing this completes our first task of writing $\langle N'(0) \rangle_\mu$ as a sum of terms.

We would like to associate each one of these terms with one element in the probability space and assign the value of the random variable at that element to be the term divided by the measure of the element. Unfortunately, there are many more terms in the sum than there are elements in the probability space. That is because forming the $\delta\beta$ derivative of any single term in the sum for $\langle N(\delta\beta) \rangle$ generates $2n$ terms, by application of the product rule. Thus we have to associate all $2n$ of these terms to a single element of the probability space. The random variable then has a value at a generic element of the probability space equal to the sum of those $2n$ terms (arising from the derivative) divided by the measure.

$$N'(0)(\omega) \equiv D(\omega) \cdot \left\{ \frac{1}{\langle \omega_{2n} | e^{-\beta A/s} | \omega_{2n-1} \rangle} \left. \frac{d}{d\delta\beta} \right|_{\delta\beta=0} \langle \omega_{2n} | e^{-(\beta+\delta\beta)A/s} | \omega_{2n-1} \rangle \right.$$

$$\left. + \ldots + \frac{1}{\langle \omega_1 | e^{-\beta B/s} | \omega_0 \rangle} \left. \frac{d}{d\delta\beta} \right|_{\delta\beta=0} \langle \omega_1 | e^{-(\beta+\delta\beta)B/s} | \omega_0 \rangle \right\}$$

$$(6.11)$$

Given the definition of the random variable $N'(0)$ we can finish the discussion of its simulation. Since the measure has not been changed, the usual simulation function picks elements of the probability space. The form of the definition of $N'(0)$ above shows that we need only compute the quantity in brackets for arbitrary ω in order to finish the simulation of $N'(0)$.

The random variable D is serially computable and its simulation does not necessitate storage of the element of the probability space in its entirety. We assert that, like D, $N'(0)$ is serially computable, so that the quantity in brackets may be accumulated in a temporary register. This follows from an inspection of the definition of $N'(0)$. The rightmost term in the brackets can be computed as ω_1 is generated from ω_0. Thereafter, ω_0 is not needed and may be discarded. The same argument clearly applies to all the terms.

Finally we address the issue of computing the derivatives of the matrix elements indicated in Eq. (6.11). Remember that the decomposition of H into

A and B was performed so that the matrix elements became computable. This same technique allows a computation of the derivative. Use the formula

$$\frac{d}{d\delta\beta}\bigg|_{\delta\beta=0} \langle\omega_{i+1}|e^{-(\beta+\delta\beta)A/s}|\omega_i\rangle \cong \frac{\langle\omega_{i+1}|e^{-(\beta+\delta\beta)A/s}|\omega_i\rangle - \langle\omega_{i+1}|e^{-\beta A/s}|\omega_i\rangle}{\delta\beta}$$

(6.12)

for suitably small $\delta\beta$. The matrix elements on the r. h. s. of this expression are calculated as usual. The matrix element derivatives are then tabulated before running the simulation and referenced as needed, identically with the conventional score factors.
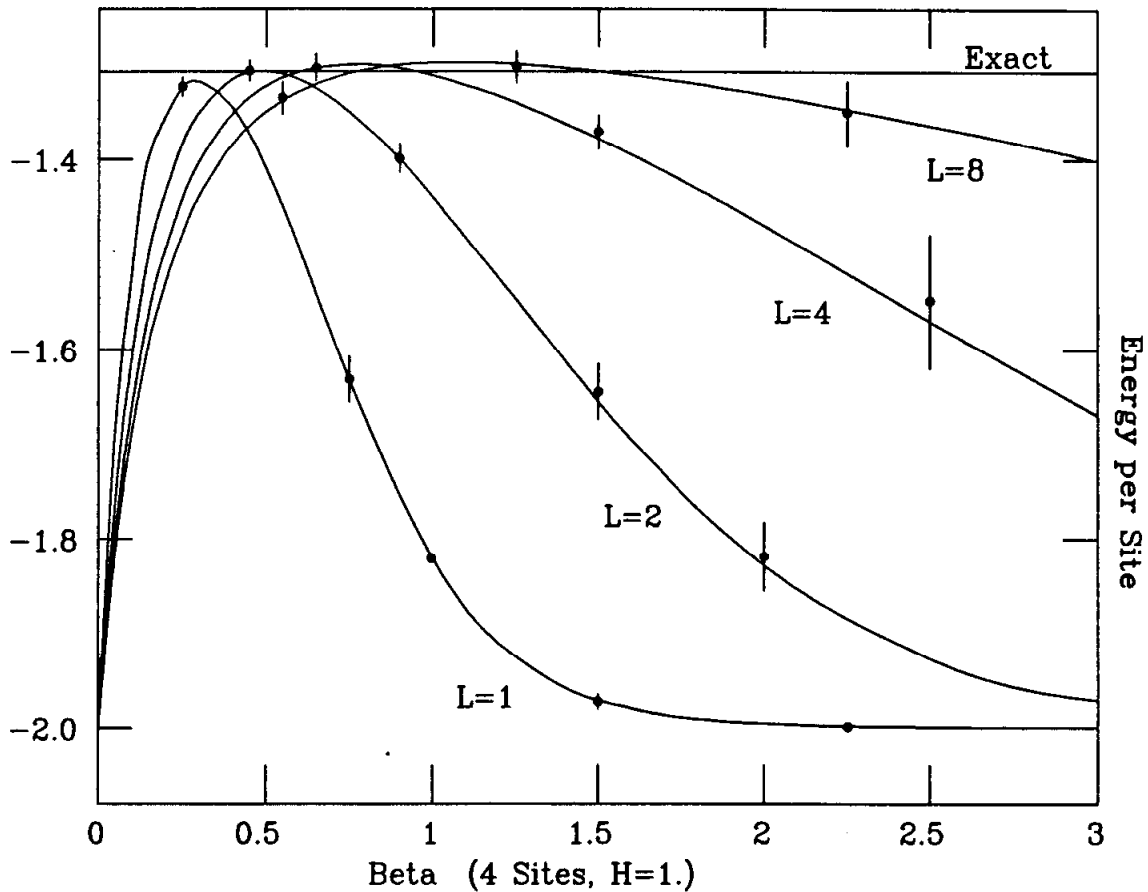
Energy Approximants vs. Beta

Figure 7. Energy approximants versus β

As with all methods presented so far for computing ground state energies,

there are systematic errors encountered in the derivative method. To appreciate the magnitude of these errors, one can undertake an exact evaluation of the ground state energy as computed by the derivative method. For models whose Hilbert space is finite dimensional and relatively small ($\lesssim 1000$), a computer can be taught to carry out this computation exactly.

In Fig. 7 results from such a calculation are presented. The curved lines are graphs of the first term on the second line of the r.h.s. of Eq. (6.8) as a function of the inverse energy β. The four curves are labelled by the value of the integer s (denoted by L in the figure) which appears in the Trotter product formula (Eq. (6.4)). The abscissa scale is energy per site of the Ising model specified in Eq. (3.1) with $N = 8$ and $h = 1$. The exact energy (obtained by fermionization and Fourier transform) is shown along with results from Projector simulations. All simulations agree (within error bars) with the appropriate analytic result. The graph underlines the significance of understanding the influence of the parameters β and s on the final answer.

6.2   TRIAL PARAMETERS

The motivation for the trial parameter technique comes from a simple analysis of the Ising model. Using the same notation as in chapter 3, compute the probability $P^{\uparrow}$ for obtaining the element of Ω with all spins up. This probability is given by Eq. (3.10) with $\omega_j(r)$ equal to 1 for all j, r. Both the Q and $P^{(A)}$ factors evaluate to 1. Each of the n factors of $P^{(B)}$ contributes p^N, where p is given by Eq. (3.24). Collecting these results, we have

$$P^{\uparrow} = e^{-\beta N h} \left[\cosh \left(\frac{\beta h}{n} \right) \right]^{nN} . \tag{6.13}$$

The score $S^{\uparrow}$ for this element is given by

$$S^{\uparrow} = \left\{ \exp \left(\frac{\beta h}{n} \right) \cdot \exp \left(\frac{\beta}{n} \right) \right\}^{Nn} = e^{\beta h N + \beta N} . \tag{6.14}$$

Now compare these expressions with the probability and score for the following

element:

$$\omega_j(r) = \begin{cases} 1 & \text{if } j \geq k \text{ and } r \neq p \text{ ; or if } j < k \\ -1 & \text{if } j \geq k \text{ and } r = p \text{ .} \end{cases} \qquad (6.15)$$

This is identical to the first element, except that a spin flip occurred at the kth time step, mediated by $P^{(B)}$. Denoting the probability and score for this element by P^{flip} and S^{flip}, one obtains

$$P^{\text{flip}}(k) = e^{-\beta N h} \left[\cosh\left(\frac{\beta h}{n}\right) \right]^{Nn-1} \left[\sinh\left(\frac{\beta h}{n}\right) \right]$$

$$S^{\text{flip}}(k) = \left[\exp\left(\frac{\beta h}{n}\right) \right]^{Nn} \left[\exp\left(\frac{\beta N}{n}\right) \right]^{k-1} \left[\exp\left(\frac{\beta(N-4)}{n}\right) \right]^{n-(k-1)}$$

$$= e^{-\beta h N - \beta N} \exp\left[-4\beta \left(1 + \frac{1-k}{n} \right) \right] . \qquad (6.16)$$

Dividing by similar factors for the spin up state, we have

$$P^{\text{flip}}(k) = \tanh\left(\frac{\beta h}{n}\right) \cdot P^{\uparrow} \qquad (6.17)$$

$$S^{\text{flip}}(k) = \exp\left[-4\beta \left(1 + \frac{1-k}{n} \right) \right] \cdot S^{\uparrow} . \qquad (6.18)$$

If $\frac{\beta h}{n} > 1$, then the probability of obtaining the flipped configuration is not appreciably different from the spin up configuration. However the factor multiplying $S^{\uparrow}$ in Eq. (6.18) may be quite small, so the score for the flipped configuration may be relatively negligible. It is not hard to understand this intuitively, and the k dependence of $S^{\text{flip}}(k)$ provides the key. If k attains its maximum value of n, then the two scores $S^{\uparrow}$ and S^{flip} are comparable, but when k is near 0 the scores can differ greatly. Every application of $e^{-\beta A/n}$ to the flipped spin basis state means a suppression factor of $e^{-4\beta/n}$. Since the flipped configuration persists until the last state in the sequence, the score is suppressed by $(n - k)$ such factors.

The following picture emerges. The measure allows the flipped spin states to occur with a probability comparable to the spin up state, although these states have negligible score in many circumstances. It is desirable to modify the measure so that trials are not wasted computing negligible contributions to the final expectation value. This modification to the measure requires compensating changes to the random variable so that the expectation value is invariant. Both these changes are easily accomplished in the trial parameter technique.

The essence of the trial parameter method is to keep the same form for the expression defining the measure, but to calculate the matrices P with a modified Hamiltonian. The change in the score random variable is exactly analogous to the change in s_i in Eq. (2.5) if p_i is changed while y_i is fixed. Choice of the proper modified Hamiltonian is suggested by the preceding analysis of the Ising model and one further requirement. The requirement is that any configuration ω for which $\mu(\omega) \cdot D(\omega)^\star$ is non-zero must have $\overline{\mu}(\omega)$ non-zero, where $\overline{\mu}$ is the new measure arising from the modified Hamiltonian. If this condition is not fulfilled, then the modified score, defined by

$$\overline{D}(\omega) \equiv \frac{\mu(\omega)}{\overline{\mu}(\omega)} \cdot D(\omega) \tag{6.19}$$

is undefined for ω. In practice this requirement is never hard to fulfill.

We now give a choice for the modified Hamiltonian in the Ising model and argue why the choice might be expected to improve convergence. Define the modified Hamiltonian $\overline{H}$ by

$$\overline{H}(\nu) \equiv -\sum_{\langle ij \rangle}^{N} \sigma_i^{(z)} \sigma_j^{(z)} - h\nu \sum_{i=1}^{N} \sigma_i^{(x)} . \tag{6.20}$$

As stated above, the measure for the probability space is calculated in the usual manner, except that $\overline{H}$ is used instead of H. The trial parameter ν has been

---

$\star$ $D(\omega)$ is any random variable similar to the one defined in Eq. (3.28).

introduced. To convince the reader that $\overline{H}$ is a good choice, we calculate the modified scores for the two configurations used above. From Eqs. (6.17,18,19):

$$\overline{S}^{\uparrow} = \frac{P^{\uparrow} \cdot S^{\uparrow}}{\overline{P}^{\uparrow}}$$

$$\overline{S}^{\text{flip}}(k) = \frac{P^{\text{flip}}(k) \cdot S^{\text{flip}}(k)}{\overline{P}^{\text{flip}}(k)} \tag{6.21}$$

$$= \tanh\left(\frac{\beta h}{n}\right) \cdot P^{\uparrow} \cdot \exp\left[-4\beta\left(1 + \frac{1-k}{n}\right)\right] \cdot S^{\uparrow} \Big/ \tan\left(\frac{\beta h\nu}{n}\right) \cdot \overline{P}^{\uparrow} \ .$$

After incorporating the trial parameter ν, the probability and score of the flipped configuration are related to the same quantities for the spin up configuration as follows:

$$\overline{P}^{\text{flip}}(k) = \tanh\left(\frac{\beta h\nu}{n}\right) \cdot \overline{P}^{\uparrow} \tag{6.22}$$

$$\overline{S}^{\text{flip}}(k) = \exp\left[-4\beta\left(1 + \frac{1-k}{n}\right)\right] \cdot \frac{\tanh(\beta h/n)}{\tanh(\beta h\nu/n)} \cdot \overline{S}^{\uparrow} \ . \tag{6.23}$$

Before the trial parameter was introduced P^{flip} and $P^{\uparrow}$ were comparable, though S^{flip} was much smaller than $S^{\uparrow}$. We hoped to reduce P^{flip} so that the flipped configurations, which contribute little to the final expectation values, are chosen less often. By choosing ν appropriately, it is possible to adjust the modified probabilities and scores to achieve just this. If ν is a small positive number, the probability for choosing the flipped configuration is diminished while the corresponding score is scaled upwards.

In Fig. 8 we demonstrate two points: i.) the expectation of the random variable of interest is independent of ν ii.) there is an optimum value of ν for which the statistical fluctuations are smallest. The data presented there are for the four site Ising model of Eq. (3.1) with $h = 1$. The graph shows the dependence of the ground state energy as a function of the trial parameter ν,

with $\beta = .5$. The ground state energy is calculated using the derivative technique of the previous section.

For each value of ν we obtained $100,000$ samples of the random variable. These were grouped into 100 bins of 1000 samples each, and the mean from each bin was computed. The 100 means thus obtained are expected to be distributed gaussianly, according to the law of large numbers. The means and widths of these gaussian distributions are plotted as a function of ν.
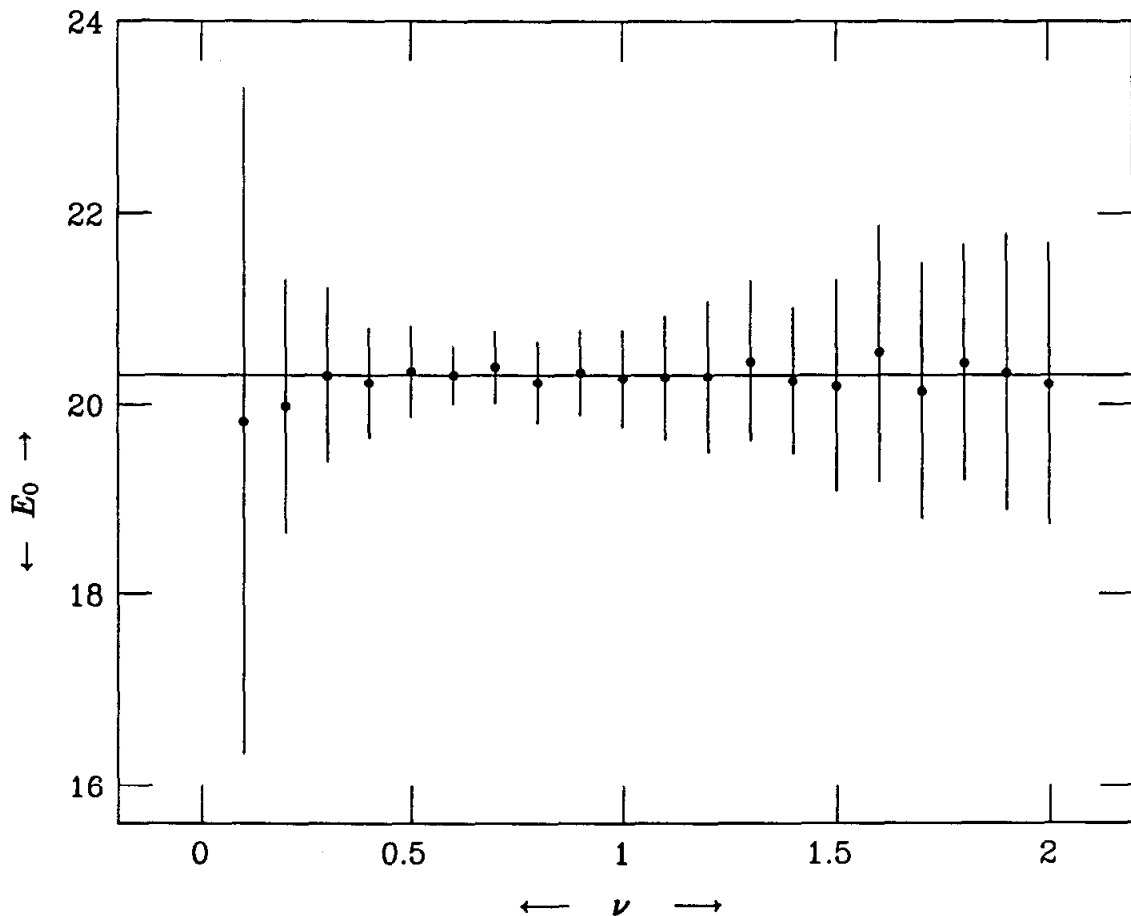

Figure 8. Dependence of energy measurements on ν

The figure shows that the ground state energy is computed correctly regardless of the value of ν, indicating that the modification of the probabilities is

correctly compensated for by the accompanying change in the scores. There is an optimum value of ν, given by choosing the gaussian distribution with the smallest width. This value of ν is somewhat less than 1, in keeping with the heuristic argument above.

Similar trial parameters are used in the PQED calculation. As in this case, the probabilities are calculated from a Hamiltonian whose form is identical to the original one, but which has different coefficients multiplying the electric and magnetic terms. The coupling constant g is replaced by a trial coupling constant denoted by g_T, and the entire Hamiltonian is multiplied by a scale factor γ. This effectively allows the coefficients for the two terms in H to be picked separately.

A heuristic argument suggests that the value of g_T should be greater than 1, and this is borne out by numerical simulations. The argument notes that when H is split into H_A and H_B, the electric term (which is responsible for suppressing transitions to states with a large amount of electric flux) has a coefficient only half as large as before. One can compensate for this inadequate suppression of transitions to states with large electric flux values by taking g_T larger than g. This use of trial parameters is responsible for much of the improved statistical behavior of our simulation as compared with De Grand et al (see reference 8).

6.3   PARALLEL CONFIGURATIONS

We now come to the multiple configuration technique. It is this technique which enhances the capability of the Projector Method for computing string energies. The multiple configuration technique is limited in applicability to certain types of simulations. We discuss the technique first heuristically, then more precisely to delimit its realm of applicability.

First consider a naive Projector Method calculation of a string energy. By string energy we mean the lowest eigenvalue in that sector of the Hilbert space which is characterized by non-vanishing eigenvalues of $\vec{\nabla} \cdot \vec{E}$ at two sites of the lattice. The eigenvalues of $\vec{\nabla} \cdot \vec{E}$ are just the values of the external charge on these

lattice sites. To see why the Projector Method calculation of the string energy is easy let us begin by considering one of the energy approximants discussed above. By inserting complete sets of energy eigenstates, the denominator of Eq. (6.1) can be rewritten

$$\langle \chi | e^{-\beta H} | \phi \rangle = \sum_n \langle \chi | n \rangle \langle n | \phi \rangle \, e^{-\beta \epsilon_n} \; . \tag{6.24}$$

The operators $\vec{\nabla} \cdot \vec{E}$ all commute with H, so the energy eigenstates $|n\rangle$ each lie in a particular external charge sector. This means that many of the terms in the above sum vanish if $|\phi\rangle$ has fixed external charge. In fact if $|\phi\rangle$ lies in the external charge sector which corresponds to the string, then

$$\lim_{\beta \to \infty} \langle \chi | e^{-\beta H} | \phi \rangle = \langle \chi | 0_{\text{str.}} \rangle \langle 0_{\text{str.}} | \phi \rangle \, e^{-\beta \epsilon_0 (\text{str.})} \tag{6.25}$$

where $|0_{\text{str.}}\rangle$ is the ground state in the string sector. An analogous argument applies to the numerator of Eq. (6.1). As a result, if we perform a standard Projector Method simulation of Eq. (6.1) with the exception that $|\phi\rangle$ is chosen to lie in the desired external charge sector, then the ground state energy extracted is the string energy. This completes our discussion of the naive Projector Method calculation of a string energy.

There is a heuristic motivation for the multiple configuration technique. Imagine simulating the ground state energy in the vanishing external charge sector. If we pick any basis state ω_i occurring in any sequence ω chosen in such a simulation, then by our above argument ω_i surely lies in the vanishing external charge sector. It is possible to invent an operator which, when applied to such an ω_i, converts it into a new state $\bar{\omega}_i$ which is in some non-vanishing external charge sector. We will exhibit such an operator shortly. At any rate, the existence of such an operator is very suggestive. It indicates that while simulating the ground state energy of the PQED system, we can apply the above operator to each basis state generated and thus simultaneously generate elements from the part of the probability space relevant to the fixed external charge problem. Of course the

states generated in this fashion will occur with a probability not equal to their measure, but this problem can be fixed in the scores.

Let us be more precise. Consider the operator

$$\mathcal{O} \equiv \exp\left(i\sum_j S_{L_j}\, \theta_{L_j}\right) \tag{6.26}$$

where each S_{L_j} assumes integer values only. Its commutator with the lattice version of the electric field divergence is easy to compute.

$$[G_{V_i},\ \mathcal{O}] = \sum_j d_1(j,i)\, S_{L_j} \cdot \mathcal{O} \tag{6.27}$$

Now if $|n\rangle$ is a basis state whose eigenvalues under all G_{V_i} are 0, observe the consequence of applying the above equality:

$$G_{V_i}\,|n\rangle = 0\,|n\rangle$$

$$G_{V_i}(\mathcal{O}\,|n\rangle) = ([G_{V_i},\ \mathcal{O}] + \mathcal{O}G_{V_i})\,|n\rangle \tag{6.28}$$

$$= \sum_j d_1(j,i)\, S_{L_j} \cdot \mathcal{O}\,|n\rangle \ .$$

Thus the state $\mathcal{O}\,|n\rangle$ has fixed external charge $\sum_j d_1(j,i)S_{L_j}$. The operator $\mathcal{O}$ may be thought of as creating a configuration of external charges on the lattice whose values at the lattice sites equal the divergence of the discretized vector field S.

A point crucial to the success of the multiple configuration technique is the following. Imagine applying the operator $\mathcal{O}$ to the entire vanishing external charge sector of the PQED Hilbert space. The result is exactly the entire sector of the Hilbert space with external charge $\sum_j d_1(j,i)S_{L_j}$. Furthermore, the mapping of these sectors one to the other is an isomorphism: the null space of $\mathcal{O}$ is trivial and the map is surjective. The necessity for these conditions will become apparent.

In a standard simulation of a matrix element such as in Eq. (5.25), choosing $|\chi\rangle$, $|\phi\rangle$, NT and β dictates the measure on the probability space. The random variable whose expectation gives the matrix element is then defined in a natural way. In the multiple configuration technique, while such a simulation is proceeding, we will simultaneously calculate the expectation value of a matrix element identical with the first, but possessed of a different initial state, $|\phi\rangle_{\text{str.}}$. If such a simulation were done independently, it would inherit a particular measure depending on $|\chi\rangle$, $|\phi\rangle_{\text{str.}}$, NT and β. Instead of this measure, the multiple configuration dictates that we define a measure*

$$\overline{\mu}(\omega) \equiv \mu\left(\mathcal{O}^{-1}(\omega)\right) . \tag{6.29}$$

Now a random variable is defined in the natural way so that its expectation gives the desired matrix element.

$$\overline{\mathcal{D}}(\omega) \equiv \frac{\langle\chi|\omega_{2n}\rangle\langle\omega_{2n}|e^{-\beta A/s}|\omega_{2n-1}\rangle \ldots \langle\omega_1|e^{-\beta B/s}|\omega_0\rangle\langle\omega_0|\phi\rangle_{\text{str.}}}{\overline{\mu}(\omega)} . \tag{6.30}$$

By construction we have

$$\int \overline{\mathcal{D}} \, d\overline{\mu} = \langle\chi|e^{-\beta H}|\phi\rangle_{\text{str.}} . \tag{6.31}$$

It should now be clear why $\mathcal{O}$ has to satisfy the twin requirements of injectivity and surjectivity. If $\mathcal{O}$ were not injective, then we could find pairs of elements of the probability space that get mapped into the same element. In this case we would have to define $\overline{\mu}$ to be a sum of μ evaluated at several elements. Though in principle this is possible, it would be awkward at best. If $\mathcal{O}$ were not surjective, then for some ω, $\mathcal{O}^{-1}(\omega)$ would not exist. This would mean that certain elements of the probability space relevant to the string problem would never be chosen. In these circumstances it is not clear how to modify the definition of $\overline{\mathcal{D}}$ so that its expectation still yields the desired matrix element.

---

* Here we have extended the domain of $\mathcal{O}$ from elements of Hilbert space to sequences of such, according to $\left(\mathcal{O}^{-1}(\omega)\right)_i \equiv \mathcal{O}^{-1}(\omega_i)$.

An alternative viewpoint of the same mathematical formalism exposes one advantage of the multiple configuration technique. Above we imagined simulating one particular probability space using a simulation function M, a measure μ and random variable $\mathcal{D}$. Simultaneously we wished to simulate a probability space with simulation function $M^{(\text{str.})}$, measure $\mu^{(\text{str.})}$, and random variable $\mathcal{D}^{(\text{str.})}$. Instead we used for the second problem the simulation function $\mathcal{O} \cdot M$, which is certainly different from $M^{(\text{str.})}$. Additionally the measure on the space is not $\mu^{(\text{str.})}$, but $\overline{\mu}$. To compensate, we defined a new random variable $\overline{\mathcal{D}}$ on the space which satisfies

$$\int \mathcal{D}^{(\text{str.})} \, d\mu^{(\text{str.})} = \int \overline{\mathcal{D}} \, d\overline{\mu} \ . \tag{6.32}$$

The alternative viewpoint is to transfer all operations back to the original probability space. The original simulation takes place with the random variable $\mathcal{D}$. The multiple configuration part employs a random variable $\overline{\overline{\mathcal{D}}}$ defined on the same space via

$$\overline{\overline{\mathcal{D}}}(\omega) = \overline{\mathcal{D}}\left(\mathcal{O}(\omega)\right) \ . \tag{6.33}$$

Clearly,

$$\int \overline{\overline{\mathcal{D}}} \, d\mu = \int \overline{\mathcal{D}} \, d\overline{\mu} \ . \tag{6.34}$$

As in the parallel score technique, we find that the variables $\mathcal{D}$ and $\overline{\overline{\mathcal{D}}}$ are highly correlated. This correlation can be used to reduce the statistical fluctuations in calculating string tensions.

To combine the parallel score technique with the multiple configuration technique, begin by noting that the string tension is the difference in the string and vacuum energy.

$$
\begin{aligned}
T &= \epsilon_{\text{str.}} - \epsilon_0 \\
&= -\frac{1}{\delta\beta} \left(\ell n \ e^{-\delta\beta\epsilon_{\text{str.}}} - \ell n \ e^{-\delta\beta\epsilon_0} \right)
\end{aligned}
\tag{6.35}
$$

Substitute ratios of matrix elements from Eq. (6.1) for the exponentials.

$$T = -\frac{1}{\delta\beta}\left\{\ell n\left(\frac{\langle\chi|e^{-(\beta+\delta\beta)H}|\phi\rangle_{\text{str.}}}{\langle\chi|e^{-\beta H}|\phi\rangle_{\text{str.}}}\right) - \ell n\left(\frac{\langle\chi|e^{-(\beta+\delta\beta)H}|\phi\rangle}{\langle\chi|e^{-\beta H}|\phi\rangle}\right)\right\}$$

$$= -\frac{1}{\delta\beta}\,\ell n\left(\frac{\langle\chi|e^{-(\beta+\delta\beta)H}|\phi\rangle_{\text{str.}}}{\langle\chi|e^{-(\beta+\delta\beta)H}|\phi\rangle}\cdot\frac{\langle\chi|e^{-\beta H}|\phi\rangle}{\langle\chi|e^{-\beta H}|\phi\rangle_{\text{str.}}}\right)\,.$$

(6.36)

From our discussion in the parallel score section and above, it can be seen that all four matrix elements in the parentheses can be simulated at the same time. Cancelling of fluctuations makes the above quantity quite stable. In fact, strings of different lengths can be calculated simultaneously, further increasing the efficiency of our technique. Detailed numerical results of this approach are presented in reference 3.

APPENDIX A

MATHEMATICAL ASIDES

We include the following definitions of mathematical terms[9] to avoid confusion and to make this work more self-contained. By a probability space we mean a triple of objects, $\{\Omega, \Sigma, \mu\}$. Ω is a set of objects which, it is worth noting, is usually not endowed with a topology. This means that we need not worry about the "smoothness" or "continuity" of functions defined on Ω, and that in fact Ω need not even have a particular dimension.

Σ is a σ-algebra on Ω, which means for our purposes that we should think of Σ as a subset of the set of all subsets of Ω. The fact that Σ is a σ-algebra means that not just any subset of the set of all subsets of Ω will do. Specifically, Σ must be closed under the setwise operations of complementation (if $\sigma \in \Sigma$ then $\tilde{\sigma} \equiv \{\omega \in \Omega : \omega \notin \sigma\}$ is also in Σ) and countable unions (if $\sigma_n \in \Sigma \ \forall \ n \in I$ and I is a countable index set, then $\bigcup_{n \in I} \sigma_n \in \Sigma$).

In intuitive terms, we can explain the significance of these restrictions as follows. The σ-algebra Σ will be the domain of a function which we call the measure and will introduce shortly. Certain desirable properties of the measure function are lost if we allow its domain to become too large. By requiring the domain of the measure function to be a σ-algebra, it becomes an easy matter (for mathematicians) to check that the measure function has the desired properties. We may thus view the defining properties of the σ-algebra Σ as insurance that the measure function behaves correctly.

Finally, μ is the measure function. As implied above, μ is just a real valued function whose domain is the σ-algebra Σ. In other words, μ assigns a real number to certain subsets of Ω. We have dignified the function μ with the title measure, and thus expect μ to satisfy certain special properties. They are positivity ($\mu(\sigma) \geq 0$ for all $\sigma \in \Sigma$), countable additivity ($\mu\left(\bigcup_{n \in I} \sigma_n\right) = \sum_{n \in I} \mu(\sigma_n)$ if σ_n are pairwise disjoint sets from Σ and I is a countable index

set), and normalization ($\mu \left(\bigcup_{\sigma \in \Sigma} \sigma \right) = 1$). These criteria derive significance from the fact that they allow one to interpret μ as a probability function.

To see how the interpretation comes about, imagine that we have a collection of objects and an algorithm for picking an infinite number of elements of the collection in a random but consistent way. Consistent means that any particular element occurs with a fixed frequency in any sufficiently large contiguous subset of the infinite sequence of chosen elements. Random has the simple intuitive meaning that there should be no correlation between neighboring pairs of elements in the series, as well as non-neighboring pairs, triplets, etc. The probability space Ω consists of our initial collection of objects. To understand the necessity for introducing a σ-algebra in order to model the above situation, it is important to recognize the fact that the initial collection of objects may be infinite. In this case it is easy to see that the probability for obtaining any particular element of the collection may be 0 (there is no chance of picking any members of a given finite set of real numbers from the unit interval in a finite number of trials). Thus one is forced to group objects from the collection together in order to form entities that have a chance of having one of their elements chosen. In this context, the σ-algebra Σ does exactly this : those groupings of the original collection which do not occur in Σ have no chance of any of their constituents being picked. Now the role of the measure is easy to explain : the measure assigned to an element σ of Σ gives the frequency with which one expects to observe members of σ in the original infinite series.

We will frequently ignore the subtleties touched upon in the previous paragraph and refer to processes which select elements of Ω according to their measure. Of course, if Σ is identical to the power set of Ω (the set of all subsets of Ω), the terminology is precise. This is the case in both the Ising Model and PQED. In the former case Ω will be finite, while in the latter it is countably infinite.

Of course we have proceeded in reverse here. People spoke of events occurring

with particular probabilities long before the introduction of the probability space. The probability space is merely the result of the mathematicians' attempt at abstracting some formal structure from intuitive examples. We find the language of the probability space more precise, and employ it with varying degrees of success throughout the text.

A few additional bits of structure are required to make contact with actual simulation. If, as a result of some process, an element of the set Ω is chosen according to the probabilities determined by the measure μ, it is frequently the case that we are satisfied with information less comprehensive than that required to specify uniquely the chosen element of Ω. In a typical example, elements of Ω may be the Cartesian coordinates of points in a 100-dimensional space. We may be interested only in the distance of points in the region to the origin. Thus, most of the information contained in the 100 Cartesian coordinates of a sample point is irrelevant to us. To accommodate such a situation, we define random variables on probability spaces. These are simply real valued functions defined on the space Ω. They must satisfy an additional restriction which has to do with measurability but which will not concern us.

Intuitively a random variable differs from an ordinary variable in a simple way. If x is an ordinary real variable, we imagine that it can take on any real value, without prejudice as to certain ones such as 7 or 4π. A random variable is identical, except in that it has a particular probability of taking on a given value, or range of values. In practice, if x is an ordinary real variable, then the equation $0 \leq x \leq .1$ may or may not be true depending on the value of x. If x is a random variable, then the equation $0 \leq x \leq .1$ has a particular probability of being true. In fact, if r is a function from the space Ω to the reals which is a random variable, then the probability that r lies in the range (a,b) is just $\mu(r^{-1}(a,b))$.

Two features arise immediately in connection with random variables. The first is integration and the second is probability distribution. Taking up the first, we note that a random variable, being a function defined on a measure space,

is susceptible of integration. The integral against the measure μ of a random variable r is called the expectation of r, which we write $\int r\,d\mu$. This number may or may not be infinite.

Secondly we introduce the probability distribution associated with a random variable r. As mentioned earlier, it is possible, by composing the inverse function of r with the measure function μ, to associate with certain subsets of the real line a probability. In fact, for a Lebesgue measurable set u on the real line, define $\mu_R(u)$ by $\mu_R(u) = \mu(r^{-1}(u))$. The function μ_R thus assigns a real number to each Lebesgue measurable set on the real line. The function μ_R inherits all the properties of the function μ, and thus μ_R is a probability measure on the real line. We will prove no mathematical theorems using the structure introduced above. We merely find it convenient to have available this precise language for exposition.

While we are collecting together language of a mathematical nature, it seems worthwhile to mention a certain central aspect of simulation of stochastic processes. In all our simulations, we begin by defining a probability space and some random variables. Ideally we would like to be able to compute expectations of these random variables exactly. Of course this is impossible in most cases of interest, and we rely instead on an approximation. The approximation is said to "sample the probability space" according to the measure, using random numbers.

One may analyze this process into two parts, the first of which is tractable and admits many solutions for a specific problem. This first part is that dealt with centrally by the Projector Method, and in this work. This first problem is the following: one assumes that one has at one's disposal a probability space of I^n where I^n is n Cartesian products of unit intervals, and the σ-algebra is the standard one of all Lebesgue measurable sets in n dimensions. The measure is also Lebesgue, and n may be as large as necessary. It is then incumbent on one to construct some random process from this probability space such that the possible outcomes of this random process are exactly elements of the probability

space Ω we are trying to simulate.

This is a careful way of saying that our ultimate computer program needs as input some set of n uncorrelated random numbers all in the unit interval. If we averaged the result of our program over all possible sets of n random numbers, the answer should be the same as if we had actually done the integrals over the probability space. If this last constraint has been fulfilled, then we have solved the first problem. This is what we attempt to do in the text, for a particular type of measure.

The second part of this approximation is intuitively plausible, but any attempts to give a rigorous justification lead quickly into a deep and dangerous quagmire. The second part refers to the fact that we never actually average the output of the program over all n-tuples of random numbers. Worse yet, the n-tuples are far from random. They are quite deterministic, as must be the output from any algorithm that runs on a computer. Perhaps most surprising is the fact that attempts even to define what one means by a random number or sequence of numbers often result in triviality.[10] Indeed, the most robust definition of a random number yet seems to be information theoretic: If the size of the smallest computer program which can generate a given number is comparable to the size of the number itself, then the number is said to have a high degree of randomness. Of course, from this point of view, the random numbers employed in numerical simulations are the worst possible, since they are typically generated by quite small programs.

In fact the above idea leads quickly into the field of cryptoanalysis if one chooses to look at these simulation programs backwards. Normally we assume that our random numbers are random, and that we are actually simulating the probability space in question. The output of the program is then taken to characterize the probability space. If we shift our focus to the random numbers, then the numerical simulation program is clearly nothing but a deterministic function of the input random numbers. One is then certainly tempted to attribute

aberrations in the program's output to the program's ability to decipher the non-randomness of the input parameters.

APPENDIX B

ANALYSIS OF THE SINGLE PLAQUETTE

We analyze the single plaquette Hamiltonian h, given by

$$h \equiv -a \left(\frac{\partial^2}{\partial \theta_1^2} + \frac{\partial^2}{\partial \theta_2^2} + \frac{\partial^2}{\partial \theta_3^2} + \frac{\partial^2}{\partial \theta_4^2} \right) - b \cos (\theta_1 + \theta_2 - \theta_3 - \theta_4) \qquad \text{(B1)}$$

where the Hilbert space is the space of square integrable functions defined on $[0, 2\pi]^4$ and the domain of the derivative operators is as specified in section 5.1. Write a basis element as

$$\frac{1}{(2\pi)^2} \exp \left(i \sum_{i=1}^{4} n_i \, \theta_i \right) \equiv |n_i\rangle \ . \qquad \text{(B2)}$$

The matrix elements of interest are denoted by $\mathcal{F}$,

$$\langle n_i' | e^{-\Delta h} | n_i \rangle \equiv \mathcal{F} \left(n_i' ; n_i \right) \qquad \text{(B3)}$$

tensor products of which give the matrices P. Computation of $\mathcal{F}$ begins by using the Trotter Product Formula:

$$e^{A+B} = \lim_{s \to \infty} \left(e^{A/s} \, e^{B/s} \right)^{s} \ . \qquad \text{(B4)}$$

Apply this to the single plaquette Hamiltonian with

$$A = -a \left(\frac{\partial^2}{\partial \theta_1^2} + \frac{\partial^2}{\partial \theta_2^2} + \frac{\partial^2}{\partial \theta_3^2} + \frac{\partial^2}{\partial \theta_4^2} \right) \qquad \text{(B5)}$$

$$B = -b \, \cos (\theta_1 + \theta_2 - \theta_3 - \theta_4) \ . \qquad \text{(B6)}$$

Imagine that s goes to infinity by passing through integer values. We can then evaluate $\mathcal{F}(n_i'; n_i)$ by inserting complete sets of states between each operator $e^{A/s}$

and $e^{B/s}$. The sum over the intermediate states will be done numerically. First evaluate the matrix elements of $e^{A/s}$ and $e^{B/s}$.

$$\langle n_i'|e^{A/s}|n_i \rangle = \prod_{i=1}^{4} \delta_{n_i, n_i'} \cdot \exp\left(-\frac{a}{s}\sum_{i=1}^{4} n_i^2\right) \tag{B7}$$

$$\langle n_i'|e^{B/s}|n_i \rangle = \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \int_0^{2\pi} d\theta_3 \int_0^{2\pi} d\theta_4$$

$$\times \exp\left[-\frac{b}{s}\cos(\theta_1 + \theta_2 - \theta_3 - \theta_4) + i\sum_{i=1}^{4}\theta_i(n_i' - n_i)\right] \bigg/ (2\pi)^4 \tag{B8}$$

To get the last matrix element, change variables in the θ_4 integral.

$$\varphi \equiv \theta_1 + \theta_2 - \theta_3 - \theta_4 \tag{B9}$$

$$\langle n_i'|e^{B/s}|n_i \rangle = \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \int_0^{2\pi} d\theta_3 \exp\left[i\sum_{i=1}^{3}(n_i' - n_i)\theta_i\right] \bigg/ (2\pi)^4$$

$$\times \int_{\theta_1 + \theta_2 - \theta_3}^{\theta_1 + \theta_2 - \theta_3 + 2\pi} d\varphi \exp\left[-\frac{b}{s}\cos\varphi + i(n_4' - n_4)\cdot(\theta_1 + \theta_2 - \theta_3 - \varphi)\right]. \tag{B10}$$

Use the fact that the φ-integrand is a periodic function of φ with period 2π to shift the region of integration. Group the θ-dependence into the appropriate integrands.

$$\langle n_i' | e^{B/s} | n_i \rangle = \int\limits_0^{2\pi} \exp\left[i(n_1' - n_1 + n_4' - n_4)\theta_1\right] \frac{d\theta_1}{2\pi}$$

$$\cdot \int\limits_0^{2\pi} \exp\left[i(n_2' - n_2 + n_4' - n_4)\theta_2\right] \frac{d\theta_2}{2\pi} \qquad \text{(B11)}$$

$$\cdot \int\limits_0^{2\pi} \exp\left[i(n_3' - n_3 + n_4' - n_4)\theta_2\right] \frac{d\theta_3}{2\pi}$$

$$\cdot \int\limits_0^{2\pi} \exp\left[-\frac{b}{s} \cos\varphi - i(n_4' - n_4)\varphi\right] \frac{d\varphi}{2\pi} .$$

The first three integrals are just Kronecker deltas of the eigenvalues. The last integral can be done using a generating function for modified Bessel function.

$$e^{z \cos \varphi} = I_0(z) + 2 \sum_{k=1}^{\infty} I_k(z) \cos(k\varphi) . \qquad \text{(B12)}$$

Substitute this into the integral:

$$\int\limits_0^{2\pi} \exp\left[-\frac{b}{s} \cos\varphi - i(n_4' - n_4)\varphi\right] \frac{d\varphi}{2\pi} = \int\limits_0^{2\pi} \left(I_0\left(-\frac{b}{s}\right) \right.$$

$$+ 2 \sum_{k=1}^{\infty} I_k(z) \left(-\frac{b}{s}\right) \cos(k\varphi) \Bigg) \qquad \text{(B13)}$$

$$\times \exp\left[-i(n_4' - n_4)\varphi\right] \frac{d\varphi}{2\pi} .$$

Interchanging orders of summation and integration, we get

$$
= I_0\left(-\frac{b}{s}\right) \int\limits_0^{2\pi} \exp\left[-i(n'_4 - n_4)\varphi\right] \frac{d\varphi}{2\pi}
$$

$$
+ 2\sum_{k=1}^{\infty} I_k\left(-\frac{b}{s}\right) \int\limits_0^{2\pi} \exp\left[-i(n'_4 - n_4)\varphi\right] \left(\frac{e^{ik\phi} + e^{-ik\phi}}{2}\right) \frac{d\varphi}{2\pi} \tag{B14}
$$

$$
= I_0\left(-\frac{b}{s}\right) \delta_{n'_4 - n_4, 0} + 2\sum_{k=1}^{\infty} I_k\left(-\frac{b}{s}\right) \cdot \left(\frac{1}{2}\delta_{n'_4 - n_4 - k, 0} + \frac{1}{2}\delta_{n'_4 - n_4 + k, 0}\right)
$$

$$
= I_{|n'_4 - n_4|}\left(-\frac{b}{s}\right) \ .
$$

The desired matrix element is

$$
\langle n'_i | e^{B/s} | n_i \rangle = \delta_{n'_1 - n_1, -(n'_4 - n_4)} \ \delta_{n'_2 - n_2, -(n'_4 - n_4)} \ \delta_{n'_3 - n_3, +(n'_4 - n_4)} \ I_{|n'_4 - n_4|}\left(-\frac{b}{s}\right) \ . \tag{B15}
$$

The symmetries of the potential term in the single plaquette Hamiltonian are manifest in the form of this matrix element. The operator $e^{B/s}$ can add a number of quanta $-n'_4 + n_4$ to n_1, add the same number to n_2, subtract this number from n_3, and subtract this number from n_4. Whereas a potential term with less symmetry might be expected to change all quantum numbers independently, we see that the potential term here affects only one linear combination of quantum numbers.

Returning to the Trotter product formula, note that since $e^{A/s}$ changes no quantum numbers and $e^{B/s}$ changes only one combination of the four quantum numbers, $\mathcal{F}(n'_i; n_i)$ vanishes unless we have

73

$$n_1' - n_1 = \Delta n$$

$$n_2' - n_2 = \Delta n$$

$$n_3' - n_3 = -\Delta n \qquad \text{(B16)}$$

$$n_4' - n_4 = -\Delta n$$

for some integer Δn. Actually a much stronger statement than this can be made. The most general function $\mathcal{F}$ that satisfies the above constraint may still have complicated dependence on its variables when it is non-vanishing. We will see, however, that even for non-vanishing $\mathcal{F}$, the dependence on some of the variables is quite simple.

To prove this assertion, begin by rewriting the matrix element of $e^{A/s}$. Note that

$$\sum_{i=1}^{4} n_i^2 = \left[\frac{1}{2}(n_1 + n_2 - n_3 - n_4)\right]^2 + \sum_{i=1}^{2} n_i^2 - \left[\frac{1}{2}(n_1 + n_2 - n_3 - n_4)\right]^2 . \quad \text{(B17)}$$

Now evaluate the last two terms on the r.h.s. with n_i replaced by $n_i + \Delta$ for $i = 1, 2$ and $n_i - \Delta$ for $i = 3, 4$.

$$\sum_{i=1}^{4} n_i^2 - \left[\frac{1}{2}(n_1 + n_2 - n_3 - n_4)\right]^2 \longrightarrow (n_1 + \Delta)^2 + (n_2 + \Delta)^2 + (n_3 - \Delta)^2$$

$$+ (n_4 - \Delta)^2 - \frac{1}{4}[n_1 + \Delta + n_2 + \Delta - n_3 + \Delta - n_4 + \Delta]^2$$

$$= \sum_{i=1}^{4} n_i^2 + \Delta \left(2n_1 + 2n_2 - 2n_3 - 2n_4 - 2n_1 - 2n_2 + 2n_3 + 2n_4\right)$$

$$+ \Delta^2(1 + 1 + 1 + 1 - 4) - \frac{1}{4}\left(n_1 + n_2 - n_3 - n_4\right)^2$$

$$= \sum_{i=1}^{4} n_i^2 - \left[\frac{1}{2}(n_1 + n_2 - n_3 - n_4)\right]^2 \qquad \text{(B18)}$$

We have separated $\sum_{i=1}^{4} n_i^2$ into a part which is invariant under the transitions induced by the potential term and a part which is not. The advantage of this is that when we write $\mathcal{F}(n_i'; n_i)$ as a sum of products of matrix elements of $e^{A/s}$ and $e^{B/s}$ obtained by inserting complete sets of intermediate states into the Trotter product formula, it is possible to separate the overall factor from the matrix elements of $e^{A/s}$ which corresponds to the invariant part isolated above. In this way we obtain s such factors, each equal to

$$\exp\left[-\frac{a}{s}\left(\sum_{i=1}^{4} n_i^2 - \frac{1}{4}(n_1 + n_2 - n_3 - n_4)^2\right)\right] . \tag{B19}$$

In analogy with angular momentum, we then define a reduced matrix element:

$$\langle n_i'|e^{-\Delta h}|n_i\rangle = \delta_{n_1'-n_1,-n_4'+n_4}\,\delta_{n_2'-n_2,-n_4'+n_4}\,\delta_{n_3'-n_3,n_4'-n_4}$$

$$\times \exp\left[-a\left(\sum_{i=1}^{4} n_i^2 - \frac{1}{4}(n_1 + n_2 - n_3 - n_4)^2\right)\right]\langle n_i'\|e^{-\Delta h}\|n_i\rangle . \tag{B20}$$

We hope by now the reader is convinced that the dependence of the reduced matrix element on the eigenvalues may be expressed as

$$R\left(n_i'\,;\,n_i\right) \equiv \langle n_i'\|e^{-\Delta h}\|n_i\rangle = R\left(n_4' - n_4\,,\,n_1 + n_2 - n_3 - n_4\right) . \tag{B21}$$

An objection may be raised at this point that the above argument has been unnecessarily circuitous. The symmetry factors were extracted from the approximate expression for the exact matrix element before the limit in the Trotter product formula is taken. This is not a problem, since the factors extracted from the limit were independent of the s parameter. It certainly is true, however, that a careful exploitation of the symmetries of the problem at the outset would have reduced our problem to a one dimensional quantum mechanics problem. The boundary conditions that define the domain of the derivative operators become awkward to formulate if we change variables, so the above pedestrian approach was favored.

Numerical evaluation of the reduced matrix element is straightforward. Care must be taken chiefly in deciding how large s must be before the Trotter product formula has converged adequately, and also in determining a reasonable cutoff for the sum over intermediate states. We now indicate how this may be done, while insuring that there is no error in determining R greater than some preset tolerance.

For convenience, define the following notation:

$$n_i^\Delta = \begin{cases} n_i + \Delta & \text{if } i = 1, 2 \\ n_i - \Delta & \text{if } i = 3, 4 \end{cases} . \tag{B22}$$

It follows that

$$\langle n_i' | e^{-\Delta h} | n_i \rangle = \sum_{\Delta_1, \Delta_2, \ldots,} \cdots \sum_{\Delta_{s-1}} \langle n_i' | e^{A/s} e^{B/s} | n_i^{\Delta_{s-1}} \rangle \langle n_i^{\Delta_{s-1}} | e^{A/s} e^{B/s} | n_i^{\Delta_{s-2}} \rangle$$

$$\cdots \langle n_i^{\Delta_1} | e^{A/s} e^{B/s} | n_i \rangle . \tag{B23}$$

To obtain a similar expression for the reduced matrix element, assume that there is a Δ_s such that $n_i' = n_i^{\Delta_s}$ (otherwise the matrix element $\langle n_i' | e - \Delta h | n_i \rangle$ would vanish). Replace the matrix element of $e^{A/s}$ by its non-invariant part and discard the invariant part.

$$\langle n_i' \| e^{-\Delta h} \| n_i \rangle = \sum_{\Delta_1} \sum_{\Delta_2} \cdots \sum_{\Delta_{s-1}} \langle n_i' \| e^{A/s} e^{B/s} \| n_i^{\Delta_{s-1}} \rangle \ldots \langle n_i^{\Delta_1} \| e^{A/s} e^{B/s} \| n_i \rangle \tag{B24}$$

$$\langle n_i^{\Delta_{N+1}} \| e^{A/s} e^{B/s} \| n_i^{\Delta_N} \rangle = \exp\left[-\frac{a}{s} \frac{1}{4} \left(n_1 + n_2 - n_3 - n_4 + 4\Delta_{N+1} \right)^2 \right] \tag{B25}$$

$$\times I_{|\Delta_{N+1} - \Delta_N|} \left(-\frac{b}{s} \right)$$

Rewrite the part of this matrix element using the following notation:

$$v \equiv (1, 1, -1, -1) , \qquad v \cdot n = n_1 + n_2 - n_3 - n_4 \tag{B26}$$

$$\exp\left[-\frac{a}{s}\frac{1}{4}(v\cdot n-4\Delta)^2\right] = \exp\left[-\frac{a}{s}\frac{1}{4}(v\cdot n)^2\right]\exp\left[-\frac{a}{s}2(v\cdot n)\Delta\right]$$

$$\times\exp\left(-\frac{a}{s}4\Delta^2\right).$$

(B27)

Insert this rewriting of the factor occurring in each reduced matrix element in the r.h.s. of Eq. (B24) to obtain

$$\langle n_i'\|e^{-\Delta h}\|n_i\rangle \cong \exp\left[-\frac{a}{4}(v\cdot n)^2\right]\sum_{\Delta_1}\cdots\sum_{\Delta_{s-1}}\exp\left[-\frac{a}{s}2(v\cdot n)\sum_{n=1}^{s}\Delta_n\right]$$

$$\times\exp\left[-\frac{a}{s}4\sum_{n=1}^{s}\Delta_n^2\right]\cdot I_{|\Delta_s-\Delta_{s-1}|}\left(-\frac{b}{s}\right)\cdots I_{|\Delta_1-0|}\left(-\frac{b}{s}\right).$$

(B28)

Define $\Delta_0 = 0$, so that

$$\langle n_i'\|e^{-\Delta h}\|n_i\rangle \cong \exp\left[-\frac{a}{4}(v\cdot n)^2\right]$$

$$\times\sum_{\Delta_1}\cdots\sum_{\Delta_{s-1}}\exp\left[-\frac{a}{s}\left(2(v\cdot n)\sum_{n=1}^{s}\Delta_n+4\sum_{n=1}^{s}\Delta_n^2\right)\right]$$

(B29)

$$\times\prod_{n=1}^{s}I_{|\Delta_s-\Delta_{s-1}|}\left(-\frac{b}{s}\right).$$

These sums over the Δ_n should all go from $-\infty$ to ∞ over the integers and exact equality obtains when $s \to \infty$. We have to compute the sum numerically, so a finite range for the sum must be chosen and a finite value of s picked. With these modifications the expression above is precisely the one used to calculate the reduced matrix element.

We now outline the logic involved in truncating the sum. First fix s. Now set a to 0. For this special case, Eq. (B29) becomes an exact equality.

$$\langle n_i'\|e^{-\Delta h}\|n_i\rangle = \sum_{\Delta_1}\cdots\sum_{\Delta_{s-1}}\prod_{n=1}^{s-1}I_{|\Delta_s-\Delta_{s-1}|}\left(-\frac{b}{s}\right)$$

(B30)

However, since the only difference between the reduced matrix element and the

exact matrix element lies in the treatment of the operator A, which has now been set to 0, we have

$$\langle n_i'|e^{-\Delta h}|n_i\rangle = \langle n_i'|\|e^{-\Delta h}\||n_i\rangle \ . \tag{B31}$$

Finally, combining the last two equalities and using Eq. (B15),

$$I_{|\Delta_s|}(-b) = \sum_{\Delta_1}\cdots\sum_{\Delta_s} \prod_{n=1}^{s-1} I_{|\Delta_s-\Delta_{s-1}|}\left(-\frac{b}{s}\right) \ . \tag{B32}$$

It is interesting to note the strong analogy between this equation and that which expresses the finite time Green function for a non-relativistic quantum mechanics problem as a sum over products of Green functions along all paths that share the same endpoints as the finite time Green function. More succinctly, the above equation is merely an embodiment of Huygen's principle.

To return to the subject at hand, note that we have an exact expression for the finite s sums for which the range of summation has not been truncated when $a = 0$. It is easy to check the effect of truncating the sum at finite $|\Delta_n|$ on this expression. The truncation of the sums was then chosen so that its effect is smaller than some preset tolerance. Now we wish to see that this same truncation is adequate even when $a > 0$. This follows from Eq. (B29), which shows that for $a > 0$, there is a preferential weighting of terms for which Δ_n is near $-(v \cdot n)/4$. Thus, we make sure that the range of truncation is always taken to include these preferentially weighted terms. As a is increased from 0, the terms that become most important in the sum are retained. In such a manner it is possible to confidently place an upper bound on each $|\Delta_n|$ while maintaining control over the accuracy.

In choosing the value for s, the above computation was carried out for various values of s. In particular we used $s = 1,2,4,8$. The change in any given matrix element was smaller than our preset tolerance of 10^{-6} in passing from 4 to 8. All computations were done at $s = 8$.

REFERENCES

1. Michael Creutz, Phys. Rev. D20, 1915 (1979) ; Phys. Rev. D21, 2308 (1980).

2. Monte Carlo Methods in Quantum Problems, ed. Malvin Kalos, D. Reidel Publishing Company, Boston (1984).

3. D. Dahl, R. Blankenbecler, Phys. Rev. D32, 977 (1985) ; D. Kung, D. Dahl, R. Blankenbecler, J. Fulco, R. Deza, to be published in Phys. Rev. B, Sept. 1985.

4. Phase Transitions and Critical Phenomenon, vol. 5b , pp. 1-16 ; ed. C. Domb and M. S. Green, Academic Press (1976).

5. Basic Concepts of Algebraic Topology, Fred H. Croom, Springer-Verlag (1978).

6. K. G. Wilson and J Kogut, Phys. Rep. 12C, 75 (1974).

7. Sidney D. Drell, Helen R. Quinn, Benjamin Svetitsky, Marvin Weinstein, Phys. Rev. D19, 619 (1979).

8. T. A. DeGrand, J. Potvin, Phys. Rev. D30, 1285 (1984) ; Phys. Rev. D31, 871 (1985).

9. The $P(\phi)_2$ Euclidean (Quantum) Field Theory, Barry Simon, Princeton Series in Physics (1974); Markov Random Fields, Yu. A. Rozanov, Springer-Verlag (1982) ; Real Analysis, H. L. Royden, Macmillan Publishing Co., Inc. (1968).

10. The Art of Computer Programming, vol. 2, Donald E. Knuth, Addison-Wesley Publishing Company (1981). See p. 142 - 166, especially p. 163-164.