A Quantum Monte Carlo method at fixed energy

Citation: Farhi, Edward, Jeffrey Goldstone, David Gosset, and Harvey B. Meyer. "A Quantum Monte Carlo Method at Fixed Energy." Computer Physics Communications 182, no. 8 (August 2011): 1663–1673.

As Published: http://dx.doi.org/10.1016/j.cpc.2011.04.021

Publisher: Elsevier

Persistent URL: http://hdl.handle.net/1721.1/102201

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

CERN-PH-TH/2009-248
MIT-CTP 4100

A Quantum Monte Carlo Method at Fixed Energy

Edward Farhi,[1,*] Jeffrey Goldstone,[1,†] David Gosset,[1,‡] and Harvey B. Meyer[1,2,§]

[1] *Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139*

[2] *Physics Department, CERN, 1211 Geneva 23, Switzerland*

Abstract

In this paper we explore new ways to study the zero temperature limit of quantum statistical mechanics using Quantum Monte Carlo simulations. We develop a Quantum Monte Carlo method in which one fixes the ground state energy as a parameter. The Hamiltonians we consider are of the form $H = H_0 + \lambda V$ with ground state energy E. For fixed H_0 and V, one can view E as a function of λ whereas we view λ as a function of E. We fix E and define a path integral Quantum Monte Carlo method in which a path makes no reference to the times (discrete or continuous) at which transitions occur between states. For fixed E we can determine $\lambda(E)$ and other ground state properties of H.

*Electronic address: farhi@mit.edu

†Electronic address: goldston@mit.edu

‡Electronic address: dgosset@mit.edu

§Electronic address: meyerh@mit.edu

I. INTRODUCTION

Quantum Monte Carlo methods are widely used to compute properties of quantum systems using classical sampling algorithms. In this paper we develop a novel Quantum Monte Carlo method that allows one to numerically investigate ground state properties of a quantum system.

A virtue of Quantum Monte Carlo is that one is not required to manipulate vectors in the Hilbert space corresponding to the quantum system. The dimension of this Hilbert space typically grows exponentially with the physical size of the system. Instead, Quantum Monte Carlo methods map the problem of approximating the ground state energy (or some other observable) onto the problem of evaluating an expectation value with respect to a probability distribution $q(X)$ over a set of configurations C (so $X \in C$). In order to evaluate this expectation value, one can use a classical Markov chain Monte Carlo algorithm to sample configurations from the distribution q. Markov chain Monte Carlo works by defining a Markov chain on the space of configurations C. This Markov chain can be described by an update rule which tells you how to generate a new configuration of the chain from the current one. The Markov chain is constructed so that the limiting distribution is $q(X)$. One then applies some large number N_0 of iterations of the Markov chain to some initial configuration X_0. If N_0 is sufficiently large then after these iterations, the distribution of subsequent configurations will be arbitrarily close to q.

We now give a brief description of how our method is used to estimate properties of the ground state. We write the Hamiltonian as $H(\lambda) = H_0 + \lambda V$, where H_0 is diagonal in a given basis $\{|z\rangle\}$ and λV is off diagonal in this basis. (In an n spin system z is an n bit string.) In section II we outline our assumptions and restrictions on H_0 and V. With these choices, the ground state energy is always less than or equal to zero, and we will see that for each value of $E < 0$, there exists one positive value $\lambda(E)$ such that the ground state of $H(\lambda(E))$ has energy E. To use our Monte Carlo method, one first must fix $E < 0$ and a large integer m. We define a path of length m to be a sequence $\{z_1, ..., z_m\}$, where each n bit string z_i is the label of the state $|z_i\rangle$. These paths are the configurations of the previous paragraph. We will define a probability distribution $f(\{z_1, ..., z_m\})$ over the set of all paths of length m (this distribution is also a function of the value of E which was chosen). We will show how the function $\lambda(E)$ can be obtained by computing an average with respect to

the probability distribution f.

To motivate our method and to get a general idea of how it works, consider the function

$$G(E,\lambda) = Tr\left[\left(\frac{-\lambda}{H(\lambda)-E}\right)V\right]. \tag{1}$$

Assuming that $E<0$ and $\lambda>0$ are chosen so that the Taylor series expansion converges, we can write

$$\begin{aligned} G(E,\lambda) &= Tr\left[\left(\frac{-\lambda}{1+\frac{\lambda}{H_0-E}V}\right)\frac{1}{H_0-E}V\right] \\ &= \sum_{m=1}^{\infty} Tr\left[\left(\frac{-\lambda}{H_0-E}V\right)^m\right]. \end{aligned} \tag{2}$$

It is clear from the expression in equation 1 that the function $G(E,\lambda)$ blows up when $E \to E_g(\lambda)$, where $E_g(\lambda)$ is the ground state energy of $H(\lambda)$. Equivalently we can say that at a fixed value of E the blow up occurs as $\lambda \to \lambda(E)$, where $E_g(\lambda(E)) = E$. At this value of λ the Taylor series expansion must diverge. In fact, this divergence occurs because as m becomes large, terms in the series approach 1 for large m (here we have made some assumptions about the Hamiltonian which we discuss in the next section) so

$$Tr\left[\left(\frac{-\lambda(E)}{H_0-E}V\right)^m\right] \approx 1\,.$$

For the remainder of this section we assume that m is large enough to make $\approx$ close to $=$. By inserting complete sets of states in the basis that diagonalizes H_0 we can express the LHS as a sum over paths

$$(\lambda(E))^m \sum_{\{z_1,\ldots,z_m\}} \langle z_1|-V|z_m\rangle\langle z_m|-V|z_{m-1}\rangle\ldots\langle z_2|-V|z_1\rangle \prod_{i=1}^{m}\frac{1}{E_i-E} \approx 1$$

where $E_i = \langle z_i|H_0|z_i\rangle$. Now taking the log and differentiating with respect to E, we obtain

$$\begin{aligned} -\frac{1}{\lambda(E)}\frac{d\lambda(E)}{dE} &\approx \sum_{\{z_1,\ldots,z_m\}} f\left(\{z_1,\ldots,z_m\}\right)\left(\frac{1}{m}\sum_{i=1}^{m}\frac{1}{E_i-E}\right) \\ &= \langle\frac{1}{m}\sum_{i=1}^{m}\frac{1}{E_i-E}\rangle_f\,. \end{aligned} \tag{3}$$

Here the expectation value is taken with respect to the measure f on paths defined by

$$f\left(\{z_1,\ldots,z_m\}\right) = \frac{1}{F}\langle z_1|-V|z_m\rangle\langle z_m|-V|z_{m-1}\rangle\ldots\langle z_2|-V|z_1\rangle\prod_{i=1}^{m}\frac{1}{E_i-E}$$

where F is a normalizing constant. We will show how to sample with respect to the distribution f in a way that makes numerical work possible. Sampling from the distribution f will also allow us to compute $-\frac{1}{\lambda(E)}\frac{d\lambda(E)}{dE}$ from equation 3 as well as other properties of the ground state.

Our paper is organized as follows. In section II we describe the types of Hamiltonians for which our method applies. In section III we outline the new method that we propose. In section IV we explicitly construct Monte Carlo update rules for the case where $V = -\sum_{i=1}^{n}\sigma_x^i$, and we give numerical data using our algorithm at $n = 16$ where we are able to compare with exact diagonalization. In section V we review the continuous imaginary time Quantum Monte Carlo method [6], which is based on the thermal path integral. We also derive a novel estimator in this ensemble of paths for the ground state energy which becomes exact in the limit $\beta \to \infty$.

II. THE HAMILTONIAN

We consider finite dimensional Hamiltonians of the form

$$H(\lambda) = H_0 + \lambda V,$$

where H_0 is diagonal in a given basis $\{|z\rangle\}$, and V has zeros along the diagonal in this basis. We make the following assumptions about the Hamiltonian:

1. The off diagonal matrix elements of V in the basis $\{|z\rangle\}$ which diagonalizes H_0 are all either negative or zero. (This ensures that our Quantum Monte Carlo method will not suffer from a sign problem.)

2. The ground state of $H(\lambda)$ is not degenerate for any value of $\lambda \in (-\infty, \infty)$.

3. The smallest eigenvalue of H_0 is zero. Note that this condition can be fulfilled without loss of generality by adding a constant term to the Hamiltonian. Writing $|z_0\rangle \in \{|z\rangle\}$ for the unique state with $H_0|z_0\rangle = 0$, we further require that $V|z_0\rangle \neq 0$. (This implies that $|z_0\rangle$ is not an eigenvector of V since $\langle z_0|V|z_0\rangle = 0$ follows from assumption 1 above.)

We write $|\psi_g(\lambda)\rangle$ and $E_g(\lambda)$ for the ground state eigenvector and ground state energy of $H(\lambda)$. From second order perturbation theory in λ, we have that

$$\begin{aligned} \frac{d^2E_g}{d\lambda^2} &= -2\sum_{z\neq z_0}\frac{|\langle z|V|z_0\rangle|^2}{\langle z|H_0|z\rangle} \\ &< 0 \end{aligned} \tag{4}$$

where the inequality is strict because $V|z_0\rangle \neq 0$.

Using the fact that

$$\frac{dE_g}{d\lambda} = \langle\psi_g(\lambda)|V|\psi_g(\lambda)\rangle \tag{5}$$

we show that

$$\frac{dE_g}{d\lambda} = \begin{cases} > 0 \text{ , for } \lambda<0 \\ = 0 \text{ , for } \lambda = 0 \\ < 0 \text{ , for } \lambda>0 \,. \end{cases}$$

In order to obtain the inequalities, we use the variational principle. When $\lambda > 0$, the ground state energy must be less than zero, since $|z_0\rangle$ has zero expectation value for H (and $|z_0\rangle$ is not an eigenvector of $H(\lambda)$). This, together with the fact that H_0 is positive semidefinite, implies that $\langle\psi_g|V|\psi_g\rangle < 0$. The analogous result for $\lambda < 0$ is obtained in the same way. These inequalities give a qualitative picture of the curve $E_g(\lambda)$. Starting from $E_g(0) = 0$, the curve slopes downwards as it goes out from λ=0, and approaches $-\infty$ on both sides of the origin for sufficiently large $|\lambda|$. Note that this implies that for each $E < 0$ there is one positive and one negative value of λ (call them $\lambda(E)$ and $\lambda_-(E)$ respectively) such that $E_g(\lambda(E)) = E$ and $E_g(\lambda_-(E)) = E$. Furthermore, we show in appendix A 1 that it is always the case that

$$\lambda(E) \leq |\lambda_-(E)| \,. \tag{6}$$

We refer to the case where the inequality is strict as the generic case. We illustrate the qualitative features of the curve $E_g(\lambda)$ (for the generic case) in figure 1. In the nongeneric case where equality holds at some particular value of E , then in fact equality holds at every value of E and the curve $E_g(\lambda)$ is symmetric about $\lambda = 0$.

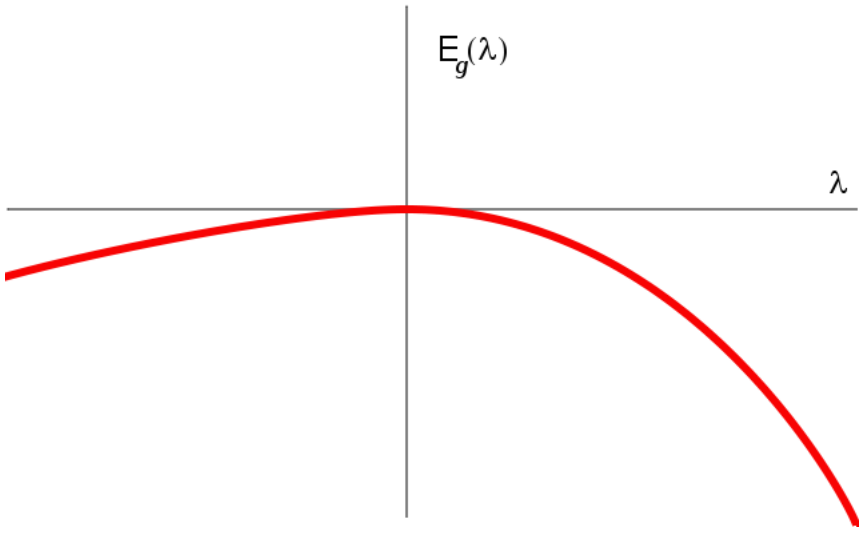

Figure 1: $E_g(\lambda)$ for the Hamiltonians we consider. As $\lambda \to \pm\infty$ we have $E_g \to -\infty$.

III. A NEW QUANTUM MONTE CARLO METHOD

Definition of the Ensemble of Paths and Relevant Estimators

As motivated in the Introduction, we now define an ensemble where the configurations are sequences $\{z_1, ..., z_m\}$ (where each z_i is an n bit string) , and we show how properties of the ground state can be computed in this ensemble. We refer to the sequences $\{z_1, ..., z_m\}$ as paths.

To begin, we fix $E < 0$ and a large integer m as parameters. As in section II, we take $\lambda(E)$ to be the positive value of λ such that $H(\lambda)$ has ground state energy E, with corresponding eigenvector $|\psi_g(\lambda(E))\rangle$. We now describe how our method allows us to approximate $\lambda(E)$ and other properties of the ground state.

Recall from the Introduction that the probability distribution f over paths is defined by

$$f(z_1, ..., z_m) \equiv \frac{1}{F(E,m)} \langle z_1| - V|z_m\rangle \ldots \langle z_2| - V|z_1\rangle \prod_{i=1}^{m} \frac{1}{E_i - E} \tag{7}$$

with[1]

$$F(E,m) \equiv \sum_{\{z_1,...,z_m\}} \langle z_1|-V|z_m\rangle ... \langle z_2|-V|z_1\rangle \prod_{i=1}^{m} \frac{1}{E_i - E} \tag{8}$$

$$= Tr\left[(\frac{-1}{H_0 - E} V)^m \right]. \tag{9}$$

As examples, we now define two quantities $\frac{\bar{\beta}(E,m)}{m}$ and $\bar{\lambda^2}(E,m)$ as ensemble averages with respect to the distribution f on paths

$$\begin{aligned} \bar{\beta}(E,m) &\equiv \sum_{\{z_1,...,z_m\}} f(\{z_1,...,z_m\}) \beta_{est}(\{z_1,...,z_m\}) = \langle \beta_{est} \rangle_f \\ \overline{\lambda^2}(E,m) &\equiv \sum_{\{z_1,...,z_m\}} f(\{z_1,...,z_m\}) \lambda^2_{est}(\{z_1,...,z_m\}) = \langle \lambda^2_{est} \rangle_f \end{aligned} \tag{10}$$

where we have defined the estimators (hence the subscript)

$$\beta_{est}(\{z_1,...,z_m\}) \equiv \sum_{i=1}^{m} \frac{1}{E_i - E} \tag{11}$$

$$\lambda^2_{est}(\{z_1,...,z_m\}) \equiv \frac{1}{m} \sum_{i=1}^{m} \delta_{z_{i+2} z_i} (E_{i+1} - E)(E_i - E) \frac{1}{\langle z_i | V^2 | z_i \rangle} \tag{12}$$

with $E_i = \langle z_i | H_0 | z_i \rangle$, and $z_{m+1} = z_1$, $z_{m+2} = z_2$. Our reason for using the symbol β_{est} will become clear in section V where we will discuss its interpretation as an inverse temperature. We show in appendix A that the ensemble averages $\frac{\bar{\beta}(E,m)}{m}$ and $\bar{\lambda^2}(E,m)$ correspond to properties of the quantum ground state in the limit $m \to \infty$

$$\lim_{m\to\infty} \frac{\bar{\beta}(E,m)}{m} = -\frac{1}{\lambda(E)} \frac{d\lambda(E)}{dE} \tag{13}$$

$$= -\frac{1}{\lambda(E)} \frac{1}{\langle \psi_g(E) | V | \psi_g(E) \rangle} \tag{14}$$

$$\lim_{m\to\infty} \overline{\lambda^2}(E,m) = (\lambda(E))^2 . \tag{15}$$

Equation 13 is equation 3 of the introduction and equation 14 follows from 5. One can also derive expressions for higher derivatives of $\log(\lambda(E))$ as averages with respect to f.

---

[1] The nongeneric case where equality holds in equation 6 can arise when there exists a unitary transformation U such that $U^\dagger V U = -V$ and $U^\dagger H_0 U = H_0$. In this case it is seen from equation 9 that $F(E,m) = 0$ when m is odd. In the nongeneric case m must always be taken to be even.

Monte Carlo Simulation

We propose to use the measure f as the basis for Monte Carlo simulations. In particular, our Monte Carlo algorithm begins by choosing $E < 0$ and a large integer m and then samples sequences of bit strings from the distribution f. One can then use the estimators from equations 10 and 12 to evaluate the quantities $\lambda(E)$ and $\langle\psi_g(\lambda(E))|V|\psi_g(\lambda(E))\rangle$, using equations 14 and 15 for the limiting behaviour of these estimators. We do not construct a general method for sampling from f, instead we leave it to the reader to construct such a method for the particular choice of V at hand. We do note that it is essential that whatever method is used conserves the total number m of transitions in the path. We now give an example of this for a generic spin system with $V = -\sum_{i=1}^{n} \sigma_x^i$.

IV. AN EXAMPLE: TRANSVERSE FIELD SPIN HAMILTONIANS

We now explicitly construct a method to compute averages with respect to the distribution f in the case where the Hilbert space is that of n spin $\frac{1}{2}$ particles, and $V = -\sum_{i=1}^{n} \sigma_x^i$. The Hamiltonian H_0 is an arbitrary diagonal matrix in the Pauli z basis for n spins. Our algorithm can be used to compute the average of any function of the path which is invariant under cyclic permutations of the path. For this choice of V, the spectrum of $H(\lambda)$ is symmetric about $\lambda = 0$ (so this corresponds to the nongeneric case where equality holds in equation 6 for all $E < 0$). With these choices, the paths which have nonzero weight (with respect to f) are periodic paths of m bit strings of length n where each string z differs from the previous one by a bit flip. We must take m to be even since each bit must flip an even number of times so that the path is periodic (and so the total number of bit flips m in the path must be even). In order to sample from these paths according to f, we construct a Markov Chain which has f as its limiting distribution (actually, our Markov Chain converges to the correct distribution over equivalence classes of paths which are only defined up to cyclic permutation–this is why we restrict ourselves to estimating quantities which are cyclically invariant). Note that with our choice of V we can write (see equation 7)

$$f(\{z_1, ..., z_m\}) = \frac{1}{F(E,m)} \prod_{i=1}^{m} \frac{1}{E_i - E}\,.$$

Our Markov chain is defined by the following update rule which describes how the configuration is changed at each step

1. Choose an integer $i \in \{1, ..., m\}$ uniformly at random.

2. Consider the bit strings z_{i-1}, z_i, z_{i+1} in the current path (where $z_{m+1} = z_1$). Suppose that z_{i-1} and z_i differ in bit $q_1 \in \{1, ..., n\}$, which we write as $z_i = z_{i-1} \oplus \hat{e}_{q_1}$. Also write $q_2 \in \{1, ..., n\}$ for the bit in which z_i and z_{i+1} differ, so $z_{i+1} = z_i \oplus \hat{e}_{q_2}$.

3. If $q_1 \neq q_2$ then propose to change the bit string z_i to the new value $\tilde{z}_i = z_{i-1} \oplus \hat{e}_{q_2}$. Accept this proposal with probability

$$P_{accept} = \min \left\{ 1, \frac{E_i - E}{\tilde{E}_i - E} \right\}$$

where $\tilde{E}_i = \langle \tilde{z}_i | H_0 | \tilde{z}_i \rangle$. This Monte Carlo move has the effect of interchanging 2 consecutive flips in the path (see figure 2).

4. If $q_1 = q_2$, then choose a new bit $q_{new} \in \{1, ..., n\}$ from the probability distribution

$$P(q_{new} = q) = \frac{1}{W} \frac{1}{E'_q - E} \tag{16}$$

where $E'_q = \langle z_{i-1} \oplus \hat{e}_q | H_0 | z_{i-1} \oplus \hat{e}_q \rangle$, and $W = \sum_{j=1}^{n} \frac{1}{E'_j - E}$. Then (with probability 1) change z_i to the new value $z_{i-1} \oplus \hat{e}_{q_{new}}$. This Monte Carlo move replaces a pair of consecutive flips which occur in the same bit with 2 new flips in a possibly different bit (see figure 3).

We show in appendix C that this algorithm can be used to estimate any quantity which is invariant under cyclic permutations of the path (note that all estimators we have discussed have this property).

Numerical Simulation with a Particular Choice of H_0

We have numerically tested our new Monte Carlo algorithm using a C++ computer program. In this section we show numerical data at 16 bits where we are able to compare results with exact numerical diagonalization. We studied the Hamiltonian with V as in the previous section, and H_0 corresponding to the combinatorial optimization problem Exact Cover which stems from our interest in quantum computation [3]

$$H(\lambda) = H_0 - \lambda \sum_{i=1}^{n} \sigma_x^i$$

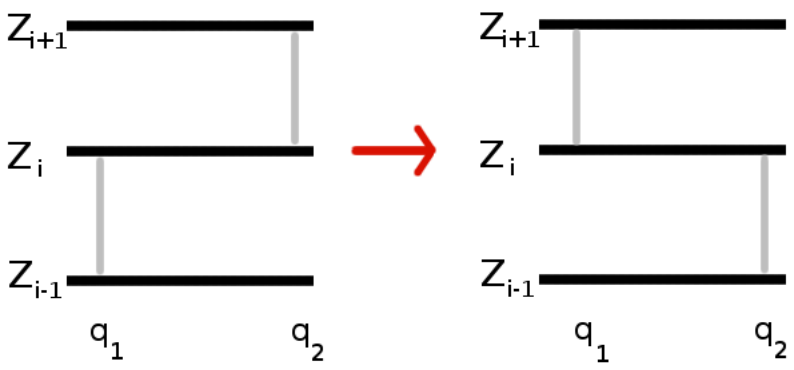

Figure 2: Monte Carlo update where the order of 2 flips in the path is interchanged

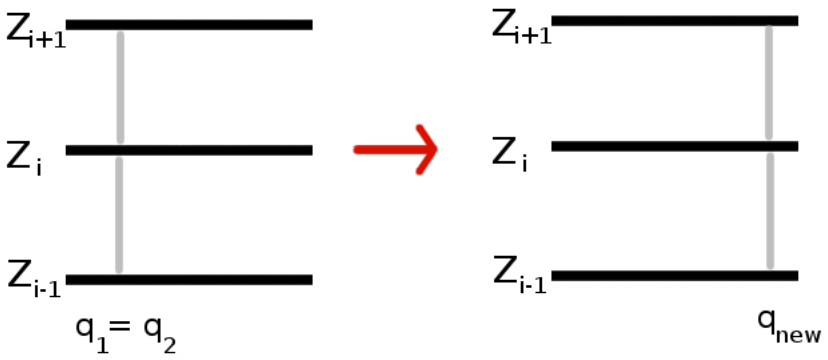

Figure 3: Monte Carlo update where 2 adjacent flips in the path which occur in the same bit are replaced by flips in a different bit.

where

$$H_0 = \sum_{c=1}^{N_c} \left(\frac{1 - \sigma_z^{i_1(c)} - \sigma_z^{i_2(c)} - \sigma_z^{i_3(c)}}{2} \right)^2 .$$

Here H_0 is a sum over N_c terms, which in computer science are called clauses. Each clause involves three distinct bits $i_1(c), i_2(c), i_3(c)$. A clause is said to be satisfied by an n bit string z if the state $|z\rangle$ has zero energy for the corresponding term in the Hamiltonian. The particular choice of N_c and the bits involved in each clause defines an instance of Exact

Cover. Such an instance is said to be satisfiable if there is an n bit string z_s which satisfies all clauses in the instance. In that case the z basis state $|z_s\rangle$ is the zero energy ground state of H_0, that is $H_0|z_s\rangle = 0$.

We generated an instance of Exact Cover on 16 bits with a unique satisfying assignment through a random procedure. Figures 4 and 5 show the values of $\lambda(E)$ and $-\lambda(E)dE(\lambda)/d\lambda$ computed using equation 10 for 200 values of E, with $m = 1000$. Statistical errors were computed using Ulli Wollf's error analysis program [10]. This data set was taken by running 10^8 Monte Carlo updates on each of two processors of a dual core laptop computer for each value of E. The two processors ran simultaneously and the total time taken for all the data was under 5 hours. We also include the curves for these quantities obtained by exact diagonalization, which are in good agreement with the Monte Carlo data. Since it is hard to see the error bars in figures 4 and 5, we have plotted the errors separately in figures 6 and 7.

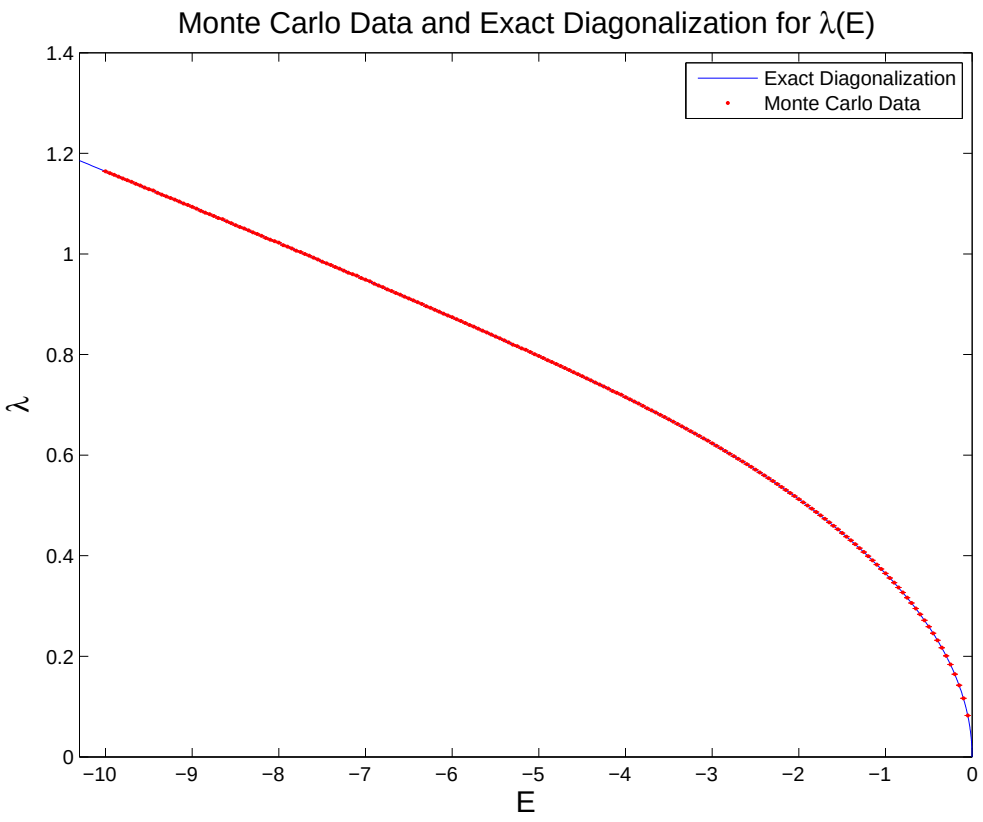

Figure 4: $\lambda(E)$ computed using Monte Carlo data and exact diagonalization for a 16 spin Hamiltonian. Statistical error bars are included for the Monte Carlo results, but they are barely visible. We have also plotted the errors separately in figure 6.

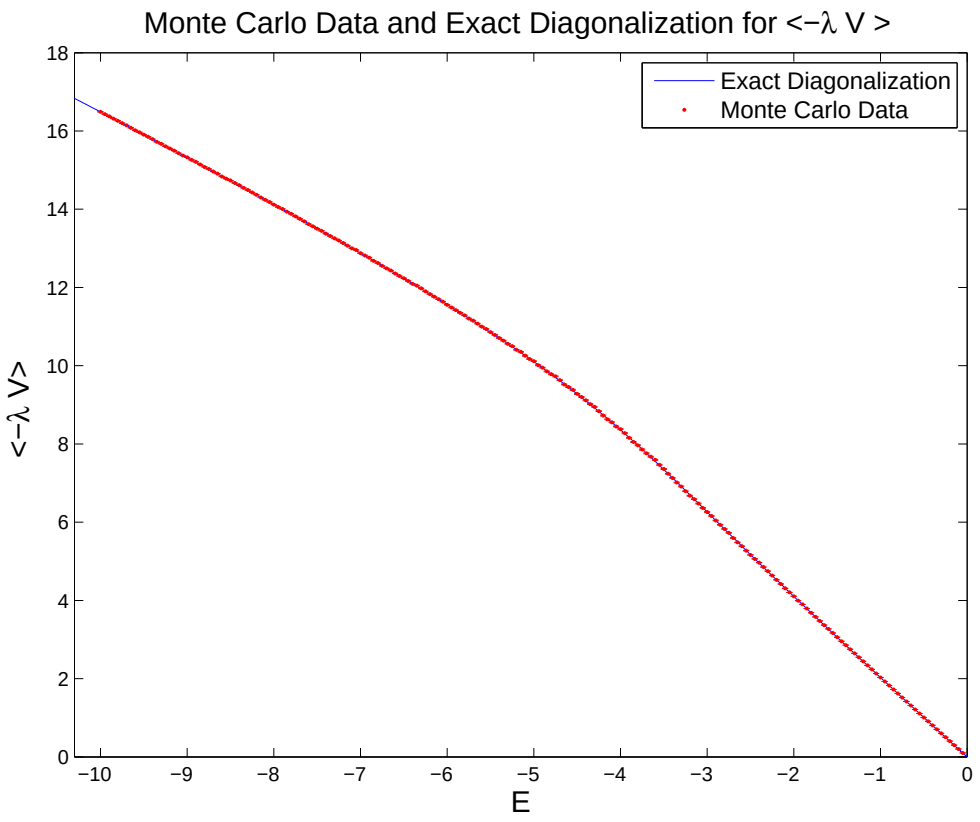

Figure 5: $-\lambda(E)\langle\psi_g(\lambda(E))|V|\psi_g(\lambda(E))\rangle$ computed using Monte Carlo data and exact diagonalization for a 16 spin Hamiltonian. Statistical error bars are included for the Monte Carlo results, but they are barely visible. Errors are also plotted separately in figure 7.

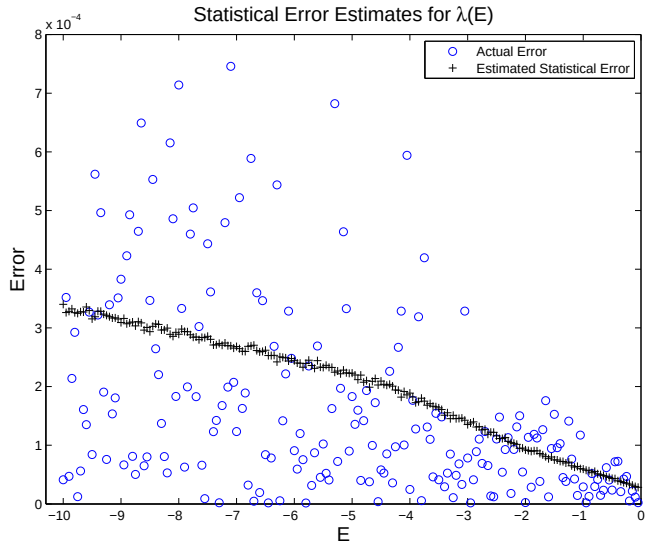

Figure 6: $\lambda(E)$ from figure 4. The black crosses show the estimated statistical error. The blue circles show the magnitude of the difference between the Monte Carlo estimates and the result of exact numerical diagonalization.

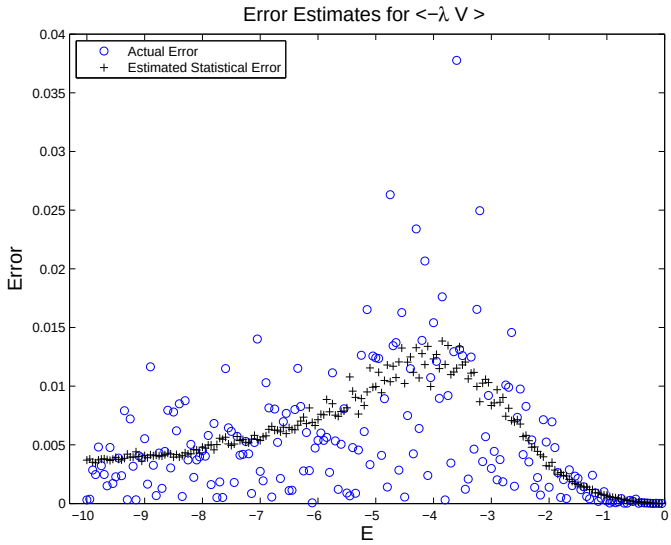

Figure 7: $-\lambda(E)\langle\psi_g(\lambda(E))|V|\psi_g(\lambda(E))\rangle$ from figure 5. The black crosses show the estimated statistical error. The blue circles show the magnitude of the difference between the Monte Carlo estimates and the result of exact numerical diagonalization.

V. A NEW ESTIMATOR FOR THE ENERGY IN STANDARD CONTINUOUS IMAGINARY TIME QUANTUM MONTE CARLO

In the ensemble of paths which was considered in the previous section, the parameters m and E are fixed and then quantities $\bar{\beta}(E,m)$ and $\overline{\lambda^2}(E,m)$ are computed as ensemble averages. In this section we review the standard approach to computing thermal averages using continuous imaginary time Quantum Monte Carlo [6]. To use this method, parameters β and λ are fixed beforehand and quantities $\bar{E}(\beta,\lambda)$ and $\bar{m}(\beta,\lambda)$ are computed as ensemble averages. We will also derive a new estimator for $\langle H\rangle = \frac{Tr[He^{-\beta H}]}{Z(\beta)}$ which is valid for large β simulations. The form of this estimator establishes a connection between the continuous imaginary time Quantum Monte Carlo method and the new method that we outlined in the previous section.

The standard method [6] is based on the expansion of the partition function

$$\begin{aligned} Z(\beta) &= Tr\left[e^{-\beta H}\right] \\ &= Tr\left[\sum_{m=0}^{\infty}(-\lambda)^m e^{-\beta H_0}\int_0^{\beta} dt_m \int_0^{t_m} dt_{m-1}... \int_0^{t2} dt_1 V_I(t_m)V_I(t_{m-1})...V_I(t_1)\right] \\ &= Tr\left[e^{-\beta H_0}\right] + \sum_{m=1}^{\infty}\left[(-\lambda)^m \sum_{\{z_1,...,z_m\}} \langle z_1|V|z_m\rangle\langle z_m|V|z_{m-1}\rangle...\langle z_2|V|z_1\rangle\right. \\ &\quad \left.\int_0^{\beta} dt_m \int_0^{t_m} dt_{m-1}... \int_0^{t_2} dt_1 e^{-(E_1 t_1 + E_2(t_2-t_1)+...+E_1(\beta-t_m))}\right] \end{aligned} \tag{17}$$

where $V_I(t) = e^{tH_0}Ve^{-tH_0}$ and $E_i = \langle z_i|H_0|z_i\rangle$. This expression is interpreted as a path integral, where a path is defined by a piecewise constant function $z(t)$ for $t \in [0,\beta]$. The function $z(t)$ takes values in the set $\{z\}$ which are the labels of the basis states $\{|z\rangle\}$ which diagonalize H_0. In the above expression for $Z(\beta)$, we have

$$z(t) = \begin{cases} z_1, & 0 \le t < t_1 \\ z_2, & t_1 \le t < t_2 \\ \vdots & \\ z_m, & t_{m-1} \le t < t_m \\ z_1, & t_m \le t \le \beta\,. \end{cases}$$

The allowed values of m are $2,3,4,5,...$ as well as $m=0$ in which case the path is constant $z(t) = z_1$ (for some z_1) for all $t \in [0,\beta]$.

We define $H_0(z(t))$ to be $\langle z(t)|H_0|z(t)\rangle$, so $\int_0^\beta H_0(z(t))dt = E_1 t_1 + E_2(t_2 - t_1) + ... + E_1(\beta - t_m)$. Equation 17 can be used to define a measure ρ on paths in imaginary time. The probability of a given path parameterized by $z(t)$ is

$$\rho(z) = \begin{cases} \frac{1}{Z(\beta)} e^{-\beta H_0(z_1)}, & m = 0 \\ \frac{1}{Z(\beta)} \lambda^m \langle z_1| - V|z_m\rangle\langle z_m| - V|z_{m-1}\rangle ... \langle z_2| - V|z_1\rangle e^{-\int_0^\beta H_0(z(t))dt} dt_1 ... dt_m, & m \neq 0 \end{cases}$$

where $Z(\beta)$ is the normalization. Our assumption that off diagonal elements of V are nonpositive guarantees that $\rho \geq 0$. The task of sampling paths from the distribution ρ can be accomplished using Markov chain Monte Carlo methods such as those outlined in references [2, 4, 6]. By using these methods to sample paths from this probability distribution, one can compute physical properties of the quantum system at an inverse temperature β. Known estimators for the expectation of the terms in the Hamiltonian are (see appendix B)

$$\langle H_0 \rangle = \frac{Tr[H_0 e^{-\beta H}]}{Z(\beta)} = \langle \frac{1}{\beta} \int_0^\beta H_0(z(t))dt \rangle_\rho \tag{18}$$

$$\langle \lambda V \rangle = \frac{Tr[\lambda V e^{-\beta H}]}{Z(\beta)} = -\langle \frac{m}{\beta} \rangle_\rho . \tag{19}$$

Here m is the number of transitions in the path $z(t)$ and is not fixed. Note that the estimator for $\langle \lambda V \rangle$ only involves the number of transitions in the path. The above two estimators can be combined to obtain $\langle H \rangle$. One can also obtain the following expressions for the variances of these estimators in the limit $\beta \to \infty$

$$\beta \mathrm{Var}\left(\frac{1}{\beta} \int_0^\beta H_0(z(t))dt \right) \overset{\beta\to\infty}{\to} -\lambda^2 \frac{d^2 E_g}{d\lambda^2}$$

$$\beta \mathrm{Var}(\frac{m}{\beta}) \overset{\beta\to\infty}{\to} -\lambda \langle \psi_g|V|\psi_g\rangle - \lambda^2 \frac{d^2 E_g}{d\lambda^2} .$$

This shows that as $\beta \to \infty$, the distributions for these estimators become sharply peaked about their mean values. We define intensive ensemble averages

$$\frac{\bar{m}(\beta,\lambda)}{\beta} \equiv \langle \frac{m}{\beta} \rangle_\rho$$

$$\bar{E}(\beta,\lambda) \equiv \langle \frac{1}{\beta} \int_0^\beta H_0(z(t))dt \rangle_\rho - \frac{\bar{m}(\beta,\lambda)}{\beta}$$

which in the $\beta \to \infty$ limit become respectively the ground state expectation value of $-\lambda V$ and the ground state energy.

A New Estimator for the Ground State Energy

We now derive the following novel estimator for the energy in the standard ensemble, which is useful in the limit $\beta \to \infty$:

$$\langle H \rangle = \langle E^\star \rangle_\rho + O\left(\frac{1}{\beta}\right)$$

where $E^\star$ is a function of the path defined to be the smallest value of E which satisfies the equation

$$\beta = \sum_{i=1}^{m+1} \frac{1}{E_i - E}\,. \tag{20}$$

In this expression the E_i are the energies of the states $|z_i\rangle$ visited along the path, with $E_{m+1} = E_1$. Note that the above equation is almost identical to equation 11 (this justifies our choice of notation β_{est}). We obtain this formula by a similar method to that used in reference [1] to obtain an alternate estimator for the energy $\langle H \rangle$. Our formula, however, is valid when some or all of the $\{E_i\}$ are the same, and therefore resolves a serious difficulty encountered in reference [1]. This may be of use in large β Monte Carlo simulations, as an alternative to the standard estimator for $\langle H \rangle$. We note in particular that this estimator does not involve the times $\{t_1, ..., t_m\}$ in the path.

Equation 20 can be derived by considering the Laplace transform of the partition function (with $s > -E_g$)

$$\int_0^\infty e^{-\beta s} Z(\beta) d\beta \;=\; Tr\left[\frac{1}{H+s}\right].$$

We can also express this as

$$\begin{aligned}\int_0^\infty e^{-\beta s} Z(\beta) d\beta \;&=\; \int_0^\infty d\beta e^{-\beta s} \sum_{m=0}^{\infty} \Bigg[\lambda^m \sum_{\{z_1,...,z_m\}} \langle z_1| - V|z_m\rangle\langle z_m| - V|z_{m-1}\rangle ... \langle z_2| - V|z_1\rangle \\ &\qquad \int_0^\infty du_{m+1} ... \int_0^\infty du_1 e^{-\sum_{i=1}^{m+1} E_i u_i} \delta(\beta - \sum_{i=1}^{m+1} u_i)\Bigg] \\ &= \sum_{m=0}^{\infty} \lambda^m \sum_{\{z_1,...,z_m\}} \langle z_1| - V|z_m\rangle ... \langle z_2| - V|z_1\rangle \prod_{i=1}^{m+1} \frac{1}{E_i + s}\,.\end{aligned}$$

Performing the inverse Laplace transform gives

$$Z(\beta) = \sum_{m=0}^{\infty} \lambda^m \sum_{\{z_1,...,z_m\}} \langle z_1| - V|z_m\rangle \ldots \langle z_2| - V|z_1\rangle \frac{1}{2\pi i} \int_C ds e^{\beta s} \left(\prod_{i=1}^{m+1} \frac{1}{E_i + s}\right).$$

C is a contour in the complex plane which encircles the poles of the integrand, which are located at $\{-E_i\}$. The expectation value of the energy can then be expressed as

$$\begin{aligned}\frac{1}{Z(\beta)}Tr[He^{-\beta H}] &= -\frac{1}{Z}\frac{dZ}{d\beta} \\ &= \left\langle \frac{\frac{1}{2\pi i}\int_C (-s)e^{\beta s}\left(\prod_{i=1}^{m+1}\frac{1}{E_i+s}\right)ds}{\frac{1}{2\pi i}\int_C e^{\beta s}\left(\prod_{i=1}^{m+1}\frac{1}{E_i+s}\right)ds}\right\rangle_\rho . \end{aligned} \tag{21}$$

The complex function $h(s) = e^{\beta s}\left(\prod_{i=1}^{m+1}\frac{1}{E_i+s}\right)$ will in general have multiple saddle points along the real axis. We can solve for the locations of these saddle points by writing

$$h(s) = e^{\beta g(s)}$$

where

$$g(s) = s - \frac{1}{\beta}\sum_{i=1}^{m+1}\log(E_i + s).$$

Saddle points occur at real values $s^\star$ where $\frac{dg}{ds}(s^\star) = 0$, which says that

$$\beta = \sum_{i=1}^{m+1}\frac{1}{E_i + s^\star}\,.$$

In the integrals in equation 21, we can choose the contour of integration along the curve of steepest descent through that saddle point $s^\star$ which is largest (it is possible to show that one can deform the contour to follow this curve without changing the encircled poles).

Performing the integrals in equation 21 and letting $E^\star = -s^\star$ we obtain

$$\frac{1}{Z(\beta)}Tr[He^{-\beta H}] = \langle E^\star\rangle + O\left(\frac{1}{\beta}\right)$$

where $E^\star$ is the smallest solution to equation 20.

VI. CONCLUSIONS

We have outlined a new approach to Quantum Monte Carlo simulations in which properties of the ground state of a quantum system are computed at a fixed value of the ground state energy. We have confirmed the validity of our method by performing a numerical simulation of a system consisting of 16 spins. Our approach involves a path integral which

does not include any jump time variables, as in the stochastic series expansion [7, 8]. We have also obtained a new estimator for the ground state energy which is valid in continuous imaginary time Quantum Monte Carlo simulations but which does not involve the imaginary time variables. We hope that our method will be applied to numerical simulations that go beyond the toy system studied in this paper.

VII. ACKNOWLEDGEMENTS

The authors thank Sam Gutmann for useful discussions. This work was supported in part by funds provided by the U.S. Department of Energy under cooperative research agreement DE-FG02-94ER40818, the W. M. Keck Foundation Center for Extreme Quantum Information Theory, the U.S Army Research Laboratory's Army Research Office through grant number W911NF-09-1-0438, the National Science Foundation through grant number CCF-0829421, and the Natural Sciences and Engineering Research Council of Canada.

---

[1] B.B. Beard. Extending monte carlo samples. *Nuclear Physics B - Proceedings Supplements*, 119:1012 – 1014, 2003. Proceedings of the XXth International Symposium on Lattice Field Theory.

[2] Edward Farhi, Jeffrey Goldstone, David Gosset, Sam Gutmann, Harvey B. Meyer, and Peter Shor. Quantum adiabatic algorithms, small gaps, and different paths, 2009. arXiv:0909.4766.

[3] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, Joshua Lapan, Andrew Lundgren, and Daniel Preda. A quantum adiabatic evolution algorithm applied to random instances of an NP-complete problem. *Science*, 292:472–475, 2001. arXiv:quant-ph/0104129.

[4] Florent Krzakala, Alberto Rosso, Guilhem Semerjian, and Francesco Zamponi. On the path integral representation for quantum spin models and its application to the quantum cavity method and to monte carlo simulations. *Physical Review B*, 78:134428, 2008.

[5] M. Newman and G. Barkema. *Monte Carlo Methods in Statistical Physics*. Clarendon Press, 1999.

[6] N. V. Prokof'ev, B. V. Svistunov, and I. S. Tupitsyn. Exact quantum monte carlo process for the statistics of discrete systems. *ZH.EKS.TEOR.FIZ.*, 64:853, 1996.

[7] A W Sandvik. A generalization of handscomb's quantum monte carlo scheme-application to the 1d hubbard model. *Journal of Physics A: Mathematical and General*, 25(13):3667–3682, 1992.

[8] A. W. Sandvik, R. R. P. Singh, and D. K. Campbell. Quantum monte carlo in the interaction representation — application to a spin-peierls model. *Physical Review B*, 56:14510, 1997.

[9] Masuo Suzuki. Relationship between d-dimensional quantal spin systems and $(d+1)$-dimensional ising systems. *Progress of Theoretical Physics*, 56(5):1454–1469, 1976.

[10] Ulli Wolff. Monte carlo errors with less errors. *Comput.Phys.Commun.156:143-153,2004; Erratum-ibid.176:383,2007.*

Appendix A: DERIVATION OF ESTIMATORS FOR THE NEW MONTE CARLO METHOD

1. Properties of the operator $A(E)$

Our analysis of the new Monte Carlo method is based on properties of the operator

$$A(E) = \left(\frac{-\lambda(E)}{H_0 - E}V\right)$$

as was seen from the Introduction.

We take $E < 0$, and $\lambda(E)$ is defined to be the positive value of λ such that the ground state of $H(\lambda)$ has energy E. We will show the following properties of this operator:

1. All eigenvalues of $A(E)$ are real and ≤ 1 in absolute value.

2. $|\psi_g(\lambda(E))\rangle$ is an eigenvector of $A(E)$ with eigenvalue $+1$. It may also be the case that $|\psi_g(-\lambda(E))\rangle$ is an eigenvector of $A(E)$ with eigenvalue -1. There are no other eigenvectors of $A(E)$ with eigenvalues ± 1.

The second property says that the subspace of states spanned by eigenvectors of $A(E)$ with ± 1 eigenvalues is either 1 or 2 dimensional. We will see that the case where it is two dimensional occurs only when $H(-\lambda(E))$ has ground state energy E.

To show that eigenvalues of $A(E)$ are real whenever $E < 0$, note that the spectrum of $A(E)$ is the same as the spectrum of the operator

$$\sqrt{H_0 - E}A(E)\frac{1}{\sqrt{H_0 - E}}$$

which is Hermitian.

Now suppose (to reach a contradiction) that $|r\rangle$ is an eigenvector of $A(E)$ with eigenvalue $R > 1$. Then

$$\left[H_0 + \lambda(E)\frac{1}{R}V\right]|r\rangle = E|r\rangle$$

which says that there exists an eigenvector of $H(\frac{\lambda(E)}{R})$ with eigenvalue E. In section II we proved that the ground state energy of $H(\lambda)$ is strictly decreasing for positive λ, which means that no eigenvector of $H(\frac{\lambda(E)}{R})$ can have energy smaller than or equal to E. This is a contradiction and so all positive eigenvalues of $A(E)$ are ≤ 1.

Now, since we have proven that all eigenvalues of $A(E)$ are ≤ 1, if it is the case that some negative eigenvalue of $A(E)$ is < -1 then the eigenvalue W of $A(E)$ which is largest in magnitude has negative sign. If this were true, then the limit

$$\lim_{k\to\infty} \frac{1}{|W|^{2k+1}} Tr\left[(A(E))^{2k+1}\right]$$

would be equal to a negative constant. But this cannot be the case since all matrix elements of $A(E)$ are positive or zero. So we have shown property 1.

We now proceed to show property 2 which was stated earlier in the appendix, and in the process we prove the inequality 6 from section II. The real nonzero eigenvalues of $A(E)$ can be related to eigenvalues of $H(\lambda)$ for some value of λ. In particular, suppose that ω is a real nonzero eigenvalue of $A(E)$ with eigenvector $|\omega\rangle$. Then

$$(H_0 - E)\,\omega|\omega\rangle \;=\; -\lambda(E)V|\omega\rangle$$

so

$$H\left(\frac{\lambda(E)}{\omega}\right)|\omega\rangle = E|\omega\rangle\,.$$

Let us write the eigenvalues of $A(E)$ as

$$1 = a_1(E) > a_2(E) \geq ... \geq a_{2^n}(E) \geq -1\,.$$

(which follows by property 1.) Then the values of λ at which $H(\lambda)$ has an eigenvalue with energy E are

$$\frac{\lambda(E)}{a_j(E)}$$

for $j \in \{1, ..., 2^n\}$. Write $\lambda_-(E)$ for the negative value of lambda such that $H(\lambda_-(E))$ has ground state energy E. The values of λ which are smallest in magnitude (and hence closest to the axis $\lambda = 0$) are $\frac{\lambda(E)}{a_1(E)} = \lambda(E) > 0$ and $\frac{\lambda(E)}{a_{2^n}(E)} = \lambda_-(E) < 0$ and these correspond to the ground state at energy E, with $|\lambda_-(E)| \geq \lambda(E)$. All other values are greater than $\lambda(E)$ in magnitude. This proves inequality 6, and also shows property 2 described above.

2. Derivation of the Estimators β_{est} and λ^2_{est}

Having derived properties 1. and 2. of the operator $A(E)$ in the previous section, we now proceed to use these properties to prove equations 14 and 15.

Our treatment below applies to both the generic case where inequality 6 is strict as well as the nongeneric case where equality holds as long as in the latter case m is always even. In either case we have

$$Tr\left[(A(E))^m\right] = (\lambda(E))^m F(E,m)\,. \tag{A1}$$

(Recall the definition of $F(E,m)$ from equation 8.) We can write this as

$$(\lambda(E))^m F(E,m) = \sum_{i=1}^{2^n} (a_i(E))^m\ .$$

Taking the log and differentiating both sides gives

$$\frac{d}{dE}\left[\log\left(\lambda(E)^m\right) + \log\left(F(E,m)\right)\right] = \frac{1}{\sum_{i=1}^{2^n}(a_i(E))^m} \sum_{j=1}^{2^n} m\,(a_j(E))^{m-1}\,\frac{da_j}{dE}\,. \tag{A2}$$

Now take the limit as $m \to \infty$. Note that in the nongeneric case the fact that m is taken to be even ensures that the denominator of equation A2 does not vanish in the limit of large (even) m. The limit of the RHS is zero for any fixed value of E. This is because in the large m limit the only terms which contribute to the sum in the numerator are those corresponding to values of j for which $|a_j(E)| = 1$. For these values of $a_j(E)$ (note j is either 1 or 2^n for these values) it is always the case that $\frac{da_j}{dE} = 0$ so these terms contribute zero to the sum.

So we have shown that

$$\lim_{m\to\infty}\left(m\frac{1}{\lambda(E)}\frac{d\lambda}{dE} + \frac{1}{F(E,m)}\frac{dF(E,m)}{dE}\right) = 0\,. \tag{A3}$$

Hence

$$\lim_{m\to\infty}\frac{1}{m}\frac{1}{F(E,m)}\frac{dF(E,m)}{dE} \;=\; -\frac{1}{\lambda(E)}\frac{d\lambda}{dE}.$$

The left hand side of this equation can be rewritten as an ensemble average, which results in

$$\lim_{m\to\infty}\frac{1}{m}\frac{1}{F(E,m)}\frac{dF(E,m)}{dE} = \lim_{m\to\infty}\langle\frac{1}{m}\sum_{i=1}^{m}\frac{1}{E_i-E}\rangle_f = -\frac{1}{\lambda(E)}\frac{d\lambda}{dE}\,. \tag{A4}$$

So for fixed m sufficiently large, one can estimate the quantity $-\frac{1}{\lambda(E)}\frac{d\lambda}{dE}$ using the ensemble average $\frac{\bar{\beta}(E,m)}{m}$. This proves equation 14.

We now derive an estimator for the quantity $\lambda(E)$ itself as an ensemble average. For this purpose we make use of the fact that

$$\lim_{m\to\infty}\frac{F(E,m-2)}{F(E,m)} = \lambda(E)^2\,.$$

Expanding the numerator and denominator as sums over paths, we obtain

$$\frac{F(E,m-2)}{F(E,m)} = \frac{\sum_{\{z_1,\ldots,z_{m-2}\}}\langle z_1|-V|z_{m-2}\rangle\ldots\langle z_2|-V|z_1\rangle\prod_{i=1}^{m-2}\frac{1}{E_i-E}}{\sum_{\{z_1,\ldots,z_m\}}\langle z_1|-V|z_m\rangle\ldots\langle z_2|-V|z_1\rangle\prod_{i=1}^{m}\frac{1}{E_i-E}}\,. \tag{A5}$$

Now rewrite the numerator as

$$\begin{aligned}F(E,m-2) \;=\; \sum_{\{z_1,\ldots,z_m\}}\Bigg(&\langle z_1|-V|z_m\rangle\langle z_m|-V|z_{m-1}\rangle\langle z_{m-1}|-V|z_{m-2}\rangle\ldots\langle z_2|-V|z_1\rangle\\ &\prod_{i=1}^{m}\frac{1}{E_i-E}\left[\delta_{z_1 z_{m-1}}(E_m-E)(E_1-E)\frac{1}{\langle z_1|V^2|z_1\rangle}\right]\Bigg)\end{aligned}$$

where $\delta_{z_1 z_{m-1}}$ is the Kronecker delta. Since our distribution $f(z_1,\ldots,z_m)$ is invariant under cyclic permutations of the bit strings $\{z_i\}$, we can write

$$\begin{aligned}F(E,m-2) \;=\; \sum_{\{z_1,\ldots,z_m\}}\Bigg(&\langle z_1|-V|z_m\rangle\langle z_m|-V|z_{m-1}\rangle\langle z_{m-1}|-V|z_{m-2}\rangle\ldots\langle z_2|-V|z_1\rangle\\ &\prod_{j=1}^{m}\frac{1}{E_j-E}\left[\frac{1}{m}\sum_{i=1}^{m}\delta_{z_{i+2}z_i}(E_{i+1}-E)(E_i-E)\frac{1}{\langle z_i|V^2|z_i\rangle}\right]\Bigg)\end{aligned}$$

where $z_{m+1}=z_1$ and $z_{m+2}=z_2$. Inserting this formula into equation A5 gives the final expression for $\lambda(E)^2$ as an ensemble average

$$\lim_{m\to\infty}\left\langle\frac{1}{m}\sum_{i=1}^{m}\delta_{z_{i+2}z_i}(E_{i+1}-E)(E_i-E)\frac{1}{\langle z_i|V^2|z_i\rangle}\right\rangle_f = \lambda(E)^2\,. \tag{A6}$$

This proves equation 15.

Appendix B: ESTIMATORS IN CONTINUOUS IMAGINARY TIME QUANTUM MONTE CARLO

In this section we derive the known estimators for $\langle H_0 \rangle$ and $\langle \lambda V \rangle$ stated in equations 18 and 19.

Estimator for $\langle H_0 \rangle$

To derive the estimator for $\langle H_0 \rangle$ we write (using the Dyson series to expand $e^{-\beta H}$)

$$\frac{Tr[H_0 e^{-\beta H}]}{Tr[e^{-\beta H}]} = \frac{1}{Z(\beta)} Tr\left[H_0 \sum_{m=0}^{\infty} (-\lambda)^m e^{-\beta H_0} \int_0^{\beta} dt_m \int_0^{t_m} dt_{m-1} ... \int_0^{t2} dt_1 V_I(t_m) V_I(t_{m-1}) ... V_I(t_1) \right]$$

where $V_I(t) = e^{tH_0} V e^{-tH_0}$. (The $m = 0$ term in the above sum is $\frac{1}{Z(\beta)} Tr\left[e^{-\beta H_0}\right]$.) Inserting complete sets of states in the basis $\{|z\rangle\}$ which diagonalizes H_0 we obtain

$$\begin{aligned} \frac{Tr[H_0 e^{-\beta H}]}{Tr[e^{-\beta H}]} &= \frac{1}{Z(\beta)} \sum_{m=0}^{\infty} \left[(-\lambda)^m \sum_{\{z_1,...,z_m\}} \langle z_1|H_0|z_1\rangle \langle z_1|V|z_m\rangle \langle z_m|V|z_{m-1}\rangle ... \langle z_2|V|z_1\rangle \right. \\ &\quad \left. \int_0^{\beta} dt_m \int_0^{t_m} dt_{m-1} ... \int_0^{t_2} dt_1 e^{-(E_1 t_1 + E_2(t_2 - t_1) + ... + E_1(\beta - t_m))} \right] \\ &= \langle H_0(z(t=0)) \rangle_{\rho} . \end{aligned}$$

where the expectation value is with respect to the measure ρ defined in section V, and $H_0(z(t=0)) = \langle z(0)|H_0|z(0)\rangle$. Noting that the measure ρ is invariant under a translation of the path by a time $x \in [0, \beta]$ (this corresponds to the transformation t_i goes to $(t_i + x)$ mod β for $i \in 1, ..., m$ followed by a reordering of the labels i to maintain time ordering), we have that

$$\langle H_0(z(t=0)) \rangle_{\rho} = \langle H_0(z(t=x)) \rangle_{\rho} \text{ , for all x}\in[0,\beta].$$

We obtain the stated estimator for $\langle H_0 \rangle$ by averaging over all $x \in [0, \beta]$

$$\langle H_0 \rangle \equiv \frac{Tr[H_0 e^{-\beta H}]}{Tr[e^{-\beta H}]} = \langle \frac{1}{\beta} \int_0^{\beta} H_0(z(x)) dx \rangle_{\rho.} .$$

Estimator for $\langle \lambda V \rangle$

As in the previous section, we begin by expanding the operator $e^{-\beta H}$

$$\frac{Tr[\lambda V e^{-\beta H}]}{Tr[e^{-\beta H}]} = \frac{1}{Z(\beta)} Tr\left[\lambda V \sum_{m=0}^{\infty} (-\lambda)^m e^{-\beta H_0} \int_0^\beta dt_m \int_0^{t_m} dt_{m-1} ... \int_0^{t2} dt_1 V_I(t_m) V_I(t_{m-1}) ... V_I(t_1)\right]. \tag{B1}$$

For $m = 0, 1, 2...$ we have

$$\begin{aligned}
&\int_0^\beta dt_{m+1} \int_0^{t_{m+1}} dt_m ... \int_0^{t2} dt_1 V_I(t_{m+1}) V_I(t_m) ... V_I(t_1) \delta(t_1) \\
&= \int_0^\beta dt_{m+1} \int_0^{t_{m+1}} dt_m ... \int_0^{t3} dt_2 V_I(t_{m+1}) V_I(t_m) ... V_I(t_2) V \\
&= \int_0^\beta dt_m \int_0^{t_m} dt_{m-1} ... \int_0^{t2} dt_1 V_I(t_m) V_I(t_{m-1}) ... V_I(t_1) V \,.
\end{aligned}$$

Plugging this expression into B1 we obtain

$$\begin{aligned}
\frac{Tr[\lambda V e^{-\beta H}]}{Tr[e^{-\beta H}]} &= \frac{1}{Z(\beta)} Tr\Bigg[(-1) \sum_{m=0}^{\infty} (-\lambda)^{m+1} e^{-\beta H_0} \\
&\quad \int_0^\beta dt_{m+1} \int_0^{t_{m+1}} dt_m ... \int_0^{t2} dt_1 V_I(t_{m+1}) V_I(t_m) ... V_I(t_1) \delta(t_1)\Bigg] \\
&= \frac{1}{Z(\beta)} Tr\Bigg[(-1) \sum_{m=1}^{\infty} (-\lambda)^{m} e^{-\beta H_0} \\
&\quad \int_0^\beta dt_m \int_0^{t_m} dt_{m-1} ... \int_0^{t2} dt_1 V_I(t_m) V_I(t_{m-1}) ... V_I(t_1) \delta(t_1)\Bigg] \\
&= -\langle (1 - \delta_{m,0}) \delta(t_1) \rangle_\rho \\
&= -\langle (1 - \delta_{m,0}) \sum_{l=1}^{m} \delta(t_l) \rangle_\rho \quad \text{(since only } t_1 \text{ can ever be 0).}
\end{aligned}$$

In the last two lines of the above m appears inside an expectation value $\langle ... \rangle_\rho$. In this context m is considered to be a function of the path. Now we can use the fact that the measure ρ over paths is invariant under translations of the path in imaginary time to write

$$\begin{aligned}
\langle \lambda V \rangle \equiv \frac{Tr[\lambda V e^{-\beta H}]}{Tr[e^{-\beta H}]} &= -\langle (1 - \delta_{m,0}) \frac{1}{\beta} \int_0^\beta \sum_{l=1}^{m} \delta(t_l - t) dt \rangle_\rho \\
&= -\langle \frac{m}{\beta} \rangle_\rho \,.
\end{aligned}$$

So $\langle \lambda V \rangle$ is $-\frac{1}{\beta}$ times the average number of jumps in a path of length β.

Appendix C: CONVERGENCE OF THE MARKOV CHAIN

We show in this section that the Markov Chain defined in section IV can be used to estimate any quantity which is invariant under cyclic permutations of the path. In order to streamline the proof, it will be useful to define a different Markov Chain over paths which has the update rule (for some fixed $0 < p < 1$)

1. With probability p do 1 update of the Markov Chain defined in section IV.
2. With probability $1 - p$ apply a random cyclic permutation to the path by letting $\{z_1, z_2, ..., z_m\} \rightarrow \{z_j, z_{j+1}, ..., z_m, z_1, ...z_{j-1}\}$ for uniformly random $j \in \{1, ..., m\}$.

In the next two sections we show that the above Markov Chain has limiting distribution f (defined in equation 7) for any choice of the parameter $0 < p < 1$. This Markov Chain with fixed $0 < p < 1$ induces a random walk on equivalence classes of paths where an equivalence class is the set of all paths related to a given path by cyclic permutation. In this equivalence class random walk, step 2 does nothing. So the limiting distribution over equivalence classes is the same whether or not step 2 is performed. If one only estimates quantities which are invariant under cyclic permutations (note that the estimators we have discussed have this property) then one can use the algorithm with $p = 1$ (so step 2 is never performed).

To show that the above Markov Chain converges to the limiting distribution f over paths, it is sufficient to verify that the update rule constructed above satisfies the following two conditions [5]:

- *Ergodicity*: Given any two paths A and B, it is possible to reach path B by starting in path A and applying the Markov chain update rule a finite number of times.
- *Detailed Balance:* For any two paths A and B,

$$f(A)P(A \rightarrow B) = f(B)P(B \rightarrow A) \tag{C1}$$

  where $P(X \rightarrow Y)$ is the probability of transitioning to the path Y given that you start in path X and apply one step of the Markov chain.

We now show that this Markov Chain satisfies these conditions.

Ergodicity

In order to show ergodicity of the Markov Chain defined above, we first note that a path can be specified either by a list of bit strings $\{z_1, ..., z_m\}$, or by one bit string z_{start} followed by a list of bits in which flips occur $\{b_1, ..., b_m\}$, with each $b_r \in \{1, ..., n\}$.

We now show that by applying the Monte Carlo update rules illustrated in figures 2 and 3, it is possible to transform an arbitrary path $A \longleftrightarrow \{z_{start}, \{b_1, ..., b_m\}\}$ into another arbitrary path $B \longleftrightarrow \{y_{start}, \{c_1, ..., c_m\}\}$ where y_{start} and z_{start} differ by an even number of bit flips. This is sufficient to show ergodicity because any path B can be cyclically permuted into a path which starts in a state $\tilde{y}_{start}$ that differs from z_{start} by an even number of flips (and our Markov chain includes moves which cyclically permute the path). We assume here that $m \geq 4$, since we are interested in the limit of large m anyways. In order to transform path A into path B, we give the following prescription:

1. First transform z_{start} into y_{start}. To do this, note that one can move any two flips b_i and b_j so that they are just before and just after z_{start} (i.e $\tilde{b}_1 = b_j$ and $\tilde{b}_m = b_i$), by applying the flip interchange rule illustrated in figure 2. So long as you do not interchange the first and last flip in the list, the bit string z_{start} will remain unchanged. Then one can flip both of these bits in the bit string z_{start} by interchanging the two flips $\tilde{b}_1$ and $\tilde{b}_m$. This describes how to flip any two bits in z_{start}, assuming that flips in these bits occur somewhere in the path. Now suppose that you wish to flip 2 bits in z_{start} but one or both of the bits does not occur in the current list of flips in the path. In that case you must first take some pair of flips which occur in some other bit q, and then move them until they are adjacent using the flip interchange rule (without ever moving them past z_{start}). Once they are adjacent, you can replace them with a pair of flips in another bit using the flip replacement rule illustrated in figure 3 . Assuming $m \geq 4$, there will always be two pairs of flips in the path which can be replaced by flips in the two bits that you desire to change in z_{start}.

2. After z_{start} has been transformed into y_{start}, one must then make the list of flips equal to $\{c_1, ..., c_m\}$. This can be done by interchanging flips and replacing pairs of flips as described above. Since this can always be achieved without interchanging the first and last flip in the path, the bit string y_{start} will remain unchanged by this procedure.

Detailed Balance

Here we demonstrate that the Markov Chain defined above satisfies the detailed balance condition from equation C1. To show this, fix two paths A and B and consider the probability of transitioning between them in one step of the Monte Carlo update rule. This probability is zero except in the following cases

1. $A = B$. In this case detailed balance is trivially satisfied.

2. A cyclic permutation of the bit strings (which is not the identity) maps the path A into the path B. In this case $f(A) = f(B)$ and the probability of transitioning from A to B in one move of the Markov Chain is also equal to the probability of the reverse transition from B to A (this is because for every cyclic permutation which maps A to B the inverse permutation is also cyclic and maps B to A). So detailed balance is satisfied.

3. A and B are the same path except for at one location. In other words, A can be described by the sequence $\{z_1^A, ..., z_m^A\}$ and B can be described by the sequence $\{z_1^B, ..., z_m^B\}$ where the corresponding bit strings are all the same except for one pair z_i^A and z_i^B. Write q_1 for the bit in which z_i^A and z_{i-1}^A differ, and q_2 for the bit in which z_i^A and z_{i+1}^A differ. We have to consider two cases depending on whether or not $q_1 = q_2$:

- Case 1: $q_1 \neq q_2$. Detailed balance follows in this case since the transition probabilities follow the Metropolis Monte Carlo rule

$$\frac{P(A \to B)}{P(B \to A)} = \min\left\{1, \frac{E_i^A - E}{E_i^B - E}\right\} \frac{1}{\min\left\{1, \frac{E_i^B - E}{E_i^A - E}\right\}}$$

$$\frac{f(A)}{f(B)} = \frac{E_i^B - E}{E_i^A - E} = \frac{P(B \to A)}{P(A \to B)}.$$

- Case 2: $q_1 = q_2$. In this case, from equation 16 we have

$$\frac{P(A \to B)}{P(B \to A)} = \frac{E_i^A - E}{E_i^B - E}$$

and

$$\frac{f(A)}{f(B)} = \frac{E_i^B - E}{E_i^A - E} = \frac{P(B \to A)}{P(A \to B)}.$$