Wang-Landau method for calculating Rényi entropies in finite-temperature quantum Monte Carlo simulations

Stephen Inglis[*]

Department of Physics and Astronomy, University of Waterloo, 200 University Avenue, Ontario, Canada, N2L 3G1

Roger G. Melko

Department of Physics and Astronomy, University of Waterloo, 200 University Avenue, Ontario, Canada, N2L 3G1 and
Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5

We implement a Wang-Landau sampling technique in quantum Monte Carlo (QMC) simulations for the purpose of calculating the Rényi entanglement entropies and associated mutual information. The algorithm converges an estimate for an analog to the density of states for stochastic series expansion QMC, allowing a direct calculation of Rényi entropies without explicit thermodynamic integration. We benchmark results for the mutual information on two-dimensional (2D) isotropic and anisotropic Heisenberg models, a 2D transverse field Ising model, and a three-dimensional Heisenberg model, confirming a critical scaling of the mutual information in cases with a finite-temperature transition. We discuss the benefits and limitations of broad sampling techniques compared to standard importance sampling methods.

I. INTRODUCTION

Entanglement is a measure of the ties that bind quantum mechanical particles across space and time. In a quantum many-body wave function, the amount of entanglement between two subregions of a system, A and B, is often quantified through a functional weighting of the eigenvalues λ_i of the reduced density matrix $\rho_A = \mathrm{Tr}_B \rho$, where there is always some set of $|\phi_i\rangle$ such that $\rho = \sum_i \lambda_i |\phi_i\rangle\langle\phi_i|$. This is used to define the so-called Rényi entanglement entropies [1],

$$S_\alpha(A) = \frac{1}{1-\alpha} \ln \left[\mathrm{Tr}\left(\rho_A^\alpha \right) \right], \qquad (1)$$

where the limit $\alpha = 1$ is the familiar von Neumann entanglement entropy. The scaling of S_1 in particular has been used to study and classify the ground-state phases and phase transitions of condensed matter systems [2–7]. Until recently it could only be accessed in numerical simulations through complete knowledge of the ground-state wave function— restricting its study to Hamiltonians that can be solved by exact diagonalization or related methods (e.g., density matrix renormalization group) [8].

In 2010, Hastings *et al.* introduced a procedure to calculate Rényi entropies for integers $\alpha \geqslant 2$ in zero-temperature projector quantum Monte Carlo (QMC) simulations through the expectation value of a *swap* operator [9]. This operator acts on two independent copies of the QMC-sampled configuration, literally "swapping" basis states in region A between the two copies. Since introduction of the algorithm, there has been a flurry of activity in calculating Rényi entropies numerically using swap-related techniques [10–14]. Soon thereafter [15], it also was demonstrated that Rényi entropies could be calculated in finite-temperature QMC simulation through adaptation of a well known "replica" trick [16], whereby the $(d+1)$-dimensional QMC simulation cell is doubled (in the case

of S_2) in size in its imaginary time direction [17,18]. The Rényi entropy is then related to the logarithm of the ratio of the partition function of the replicated system ($Z[A,2,T]$) to the partition function of the original system, $Z(T)$. Unlike the swap method that gives a Rényi entropy directly at $T = 0$, the replica trick calculated for two simulations (one for $Z[A,2,T_0]$ and one for $Z(T_0)$) alone does *not* give the Rényi entropy at the temperature T_0. Typically, a careful integration over many simulations from $T = \infty$ to $T = T_0$ must be performed to calculate the Rényi entropy [15,19]. Other approaches are possible, including a recent idea that allows direct calculation of the ratio of two partition functions at a fixed temperature without integration [20], which effectively implements ideas of the *swap* operator at finite temperature. Ultimately, such ideas should be wed into the application of broad histogram techniques, which are natural solutions for Monte Carlo simulations which require calculating the partition function.

In this paper, we use an advanced Wang-Landau [21] sampling technique for calculating the Rényi entropy in finite-temperature QMC simulation that avoids the need for many separate simulations and allows access to the Rényi entropy at any arbitrary temperature between $T = \infty$ and some chosen cutoff temperature. We outline the technique for the specific case of S_2 using a stochastic series expansion (SSE) QMC simulation. We then demonstrate the method by looking at the finite-size scaling of the Rényi mutual information in three variants of the spin-$1/2$ XXZ model in two and three dimensions and the two-dimensional (2D) transverse field Ising model.

II. METHODS

In this section we discuss the specific implementation details of the Wang-Landau method in the context of QMC simulation and calculating the mutual information between two subregions A and B. This particular implementation is guided by the established SSE QMC simulation [22–24]

*dartonias@gmail.com

and recent extensions using a modified simulation topology [15]. For adding Wang-Landau sampling to the SSE QMC simulation, we find the method outlined by Troyer *et al.* [25] is a simple and concise starting point, although it has similar limitations that have been found also to exist in classical Wang-Landau methods [26]. Going beyond Wang-Landau approaches to more general broad histogram sampling, we implement a more advanced technique detailed by Wessel *et al.* [27] that addresses some of the shortcomings with the original Wang-Landau approach but still provides the density of states needed to calculate thermodynamic properties (including the Rényi entropy) over a large range of temperatures.

A. Stochastic series expansion

The SSE [23,28] is a framework similar conceptually to world line methods [29–31] in which one simulates a d-dimensional quantum system using a $(d + 1)$-dimensional classical simulation. To begin, we recall the standard SSE implementation, which represents the partition function as

$$Z = \text{Tr}[e^{-\beta H}] = \sum_\psi \langle\psi| \sum_{n=0}^\infty \frac{(-\beta)^n H^n}{n!} |\psi\rangle, \qquad (2)$$

where $\sum_\psi$ represents the trace over some complete basis of states for the system. We rewrite the above by collecting all the prefactors to the left and inserting the resolution of the identity between each power of the Hamiltonian. Since the Hamiltonian is a sum, we can group operators into pieces that act on certain groups of sites and by whether they are diagonal or off diagonal in our basis of choice. These grouped pieces of the Hamiltonian we refer to as "bond" operators $H_{a,b}$, where a differentiates diagonal and off diagonal terms and b denotes a grouping of sites. For our formalism, $a = 1$ represents diagonal terms and $a = 2$ represents off diagonal terms, while b specifies the bond connecting two sites that the Hamiltonian piece operates on. Using this formalism we obtain the SSE representation of the partition function,

$$Z = \sum_n \frac{(-\beta)^n}{n!} \sum_{\{\psi\}} \sum_{i=0}^n \langle\psi_i|H_{a_i,b_i}|\psi_{i+1}\rangle, \qquad (3)$$

where $\sum_{\{\psi\}}$ represents the sum over all valid configurations $\{\psi_0, \ldots, \psi_{n+1}\}$ with the condition $\langle\psi_0| = \langle\psi_{n+1}|$. If we allow the sum to go from zero to infinity and include all possible lists of operators H_{a_i,b_i} for each n, this formulation is exact. In practice, contributions above some large value N are numerically insignificant and can be ignored. The cutoff N is always larger than any visited n in the importance sampling SSE QMC simulation and, as such, does not affect the statistics except to allow us to represent the set of operators as a finite list.

The two updates we use to efficiently sample only nonzero weights are referred to as the *diagonal update* and the *directed loop update* [23]. In the diagonal update we insert and remove diagonal operators from the list, changing n and the total weight, but not changing $\langle\psi_0|$. We use the change in weight at each insertion or removal and the usual Boltzmann condition to accept or reject these proposed moves. In the directed loop updates we connect all the Hamiltonian elements by what sites they interact with at each insertion of the identity in
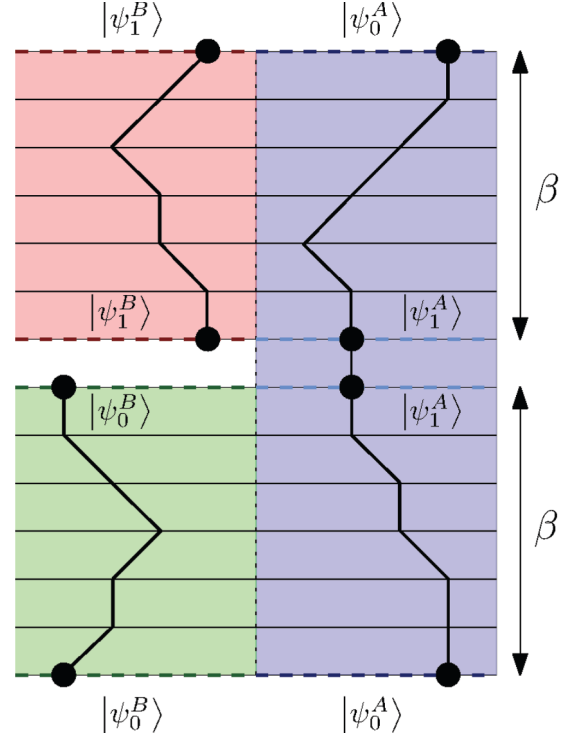

FIG. 1. (Color online) Boundary conditions of the modified simulation, related to Eq. (6). States with the same labels represent where the periodic boundary conditions of the simulation cell are, while lines represent the paths of up spins (or bosons, depending on context) in the imaginary time direction.

the expanded trace $\langle\psi_i|$. We then allow a loop to traverse this linked network of sites and change the operators (and spin configurations) it passes through, changing the *type* of each operator (and hence total weight) but not changing *which* sites each operator interacts with. Using these two updates our simulation is able to satisfy both ergodicity and detailed balance.

Recently, the SSE QMC simulation was adapted to allow for the calculation of Rényi entropy between two regions, A and its complement B [19]. To do this, the so-called replica trick was employed to estimate Eq. (1):

$$\text{Tr}(\rho_A^\alpha) = \frac{Z(A,\alpha,\beta)}{Z(\beta)^\alpha}, \qquad (4)$$

allowing the calculation of the Rényi entropies through

$$S_\alpha(\rho_A) = \frac{1}{1-\alpha}(\ln[Z(A,\alpha,\beta)] - \alpha \ln[Z(\beta)]). \qquad (5)$$

Here, the two partition functions represent two separate QMC simulations: $Z(\beta)$ is the "usual" partition function for the physical system under study, and $Z(A,\alpha,\beta)$ is a replicated partition function of the form

$$Z(A,\alpha,\beta) = \sum_\psi \langle\psi_0^A|\langle\psi_0^B|e^{-\beta H}|\psi_1^A\rangle|\psi_0^B\rangle$$

$$\otimes \langle\psi_1^A|\langle\psi_1^B|e^{-\beta H}|\psi_0^A\rangle|\psi_1^B\rangle, \qquad (6)$$

where $\langle\psi_i^A|$ represents the basis containing spins in A, while $\langle\psi_i^B|$ represent the spins not in region A and, as illustrated in Fig. 1, the boundary of region A is "stitched" together between the two replicas in the imaginary time direction.

It is apparent from Eq. (5) that the calculation of Rényi entropies involves a procedure very similar to calculating a difference of free energies. As is well known, neither the free energy, $\ln[Z(T)]$, nor in fact the partition function itself is accessible directly through importance sampling. Previous calculations of S_α were obtained using a thermodynamic integration of internal energies E, obtained over many separate simulations (or one "annealing" simulation) [15,19] or by simulations that sample the ratio of two distinct partition functions at a fixed temperature [20]. However, techniques involving extended ensembles have been employed for decades in the calculation of free energies and related quantities in a variety of Monte Carlo techniques. Below, we discuss the adaptation of one of these, Wang-Landau sampling [21], to the calculation of Rényi entropies in finite-temperature QMC techniques.

B. Wang-Landau sampling

In contrast to importance sampling where we sample states using the thermal weights of the system and take an unweighted average of observables, Wang-Landau sampling seeks to use the sampling process itself to determine the distribution of an unknown function. If we consider Eq. (3), the partition function can be rewritten as

$$Z = \sum_n \beta^n g(n), \qquad (7)$$

where n is the number of operators in the SSE formalism, Eq. (3). Since in the SSE the energy can be calculated using the estimator [32] $E = -\langle n \rangle / \beta$, by knowing $g(n)$ we are also able to calculate the energy and related variables like the specific heat for all temperatures [assuming complete knowledge of $g(n)$]. In addition, we are able to explicitly calculate the partition function for any temperature (up to some cutoff), something not possible when using importance sampling.

Unlike importance sampling in which we sample states with weights derived from the Boltzmann distribution, the Wang-Landau method samples using the weighting $W = g^{-1}(E)$ [or $g^{-1}(n)$ in the quantum case], ensuring that all energies are sampled with equal frequency. Since we do not know $g(E)$ *a priori* we start with an estimated density of states, typically flat, and sample the system assuming that our estimate is good. If $g(E)$ is correct for all energy (up to a scaling factor) then the histogram of energies sampled will be flat; if the histogram is larger for some energy E', then our estimate for $g(E')$ is too low for that energy, and vice versa. In this way we use the sampling histogram to iteratively update the density of states until the histogram is flat and $g(E)$ is converged.

C. Quantum Monte Carlo and Wang-Landau techniques

In this section we describe all the changes added to the SSE QMC simulation necessary for performing the quantum Wang-Landau technique. As mentioned before, previous work [25] has successfully implemented Wang-Landau sampling in SSE QMC simulations, as well as extensions of the method [27]. Our implementation mainly focuses on using the Wang-Landau method (as a proof of concept) and includes recent advances made in the classical Wang-Landau formalism [26].

The first challenge with all Wang-Landau techniques is representing the sampling function, here $g(n)$, in a tractable way. This problem of internal representation is that $g(n)$ ranges over 2000 orders of magnitude (for larger systems of typical Hamiltonians). Given computer precision, the common solution that works here is to only keep track of the natural logarithm of $g(n)$. This representation is entirely compatible with the simulation and allows for the largest amount of precision without using special data types. For the purpose of the below discussion, we let $G(n) = \ln(g(n))$.

The biggest change from the importance sampling simulation is a modification of the diagonal update [25] in which the number of operators [the n in $g(n)$] is changed. In the normal simulation, we compare the weight before and after, giving us the probability of adding an operator as

$$P = \min\left\{ \frac{\langle \psi_i | H_{1,b} | \psi_{i+1} \rangle \beta}{M - n}, 1 \right\}, \qquad (8)$$

where $H_{1,b}$ is a diagonal operator acting on bond b and n is the number of operators. When using the Wang-Landau technique this update changes to

$$P = \min\left\{ \frac{\langle \psi_i | H_{1,b} | \psi_{i+1} \rangle e^{G(n)-G(n+1)}}{M - n}, 1 \right\}, \qquad (9)$$

using the current internal estimate of $G(n)$. Then, whether we accept or reject the move, we update

$$g(n) = g_{\text{old}}(n) f, \qquad (10)$$

$$G(n) = G_{\text{old}}(n) + F, \qquad (11)$$

using the new n after the update. $F = \log(f)$ is the refinement parameter used to converge the function $G(n)$. The refinement parameter starts as some large value (in our case e, the natural log) and when the current histogram is found to be flat ($\pm 10\%$ of the mean for each bin), F is reduced by

$$F = F_{\text{old}}/2, \qquad (12)$$

and the histogram is reset. The directed loop update is carried out as before [22] (after each diagonal update), but $G(n)$ is not changed or used in this update since the number of operators does not change.

The above algorithm works well when the estimate of $G(n)$ is far from convergence, but more recent work on the classical Wang-Landau algorithm [26] has shown that a small saturated error may not converge with this method. The concept is simple: when the system is close to converged, say the maximum deviation of $G(n)$ from the true function is less than our flatness tolerance, then the simulation will assume the histogram is flat (within tolerance) after a small amount of sampling. This results in the refinement factor getting small very quickly without changing $G(n)$ significantly, causing early noise in the sampling of $G(n)$ to become frozen in. To correct for this, we keep track of the number of full diagonal updates (update over all elements of the operator list) and directed loop updates as a number t. We use the aforementioned update until we find that $F < 1/t$, after which we ignore the histograms and instead set the refinement parameter to
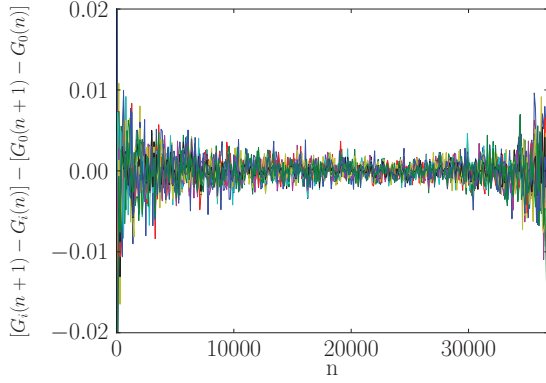
$$F = 1/t. \qquad (13)$$

FIG. 2. (Color online) Comparison of the first derivative of $G_i(n)$ using $G_0(n)$ (arbitrary) as a baseline for $i = 0, \ldots, 15$. Each color represents a different independent simulation of the 2D XXZ model of a 32×32 lattice. The noise is evenly distributed with a larger width for the minimal and maximal values of n.

The above algorithm is used for all the results we present.

D. Ensemble optimization

If we move to more general broad histogram techniques, a good extension for the SSE QMC simulation is the "minimum round trip" approach [27]. This technique tends to reduce the statistical error (shown in Fig. 2) by reshifting the sampling effort to the areas contributing most to the error, thereby being more efficient overall.

To briefly summarize the technique, it keeps track of separate histograms for upward and downward moving walkers, where the walkers are moving in the one-dimensional space of *number of operators*, n. A walker is "upward" if it most recently visited $n = 0$, and "downward" walkers are those that most recently visited $n = N$ (where N is our cutoff in the number of operators). With these two histograms we have both the total sampling (upward plus downward) and the rate of change in the fraction of upward (or downward) walkers as a function of n. The concept is that if there is a point in the simulation where the fraction of upward walkers change sharply, this is interpreted as the simulation having trouble tunneling through a barrier in configuration space. To pass through such barriers more easily, we weight the simulation to locations where the change in the fraction of walkers is large. In addition, the sampling is weighted by the inverse number of total samples—this alone would simply be Wang-Landau sampling (in that the weights would converge to the inverse density of states), but the addition of the second condition skews the weights. This process is iterated until the modified weights converge.

With the modified distribution we can sample and take the product of the histogram of visits and our skewed weights to get the density of states (in Wang-Landau the histogram of visits is flat, and so this product is not necessary in principle), and once we have this we can proceed with an identical analysis to our treatment of generating the density of states with Wang-Landau sampling alone. We implement this method for the 2D transverse field Ising model and present and discuss the results in Sec. III.

E. Calculating observables

Unlike importance sampling where the calculation of observables is aggregated over the course of the simulation, Wang-Landau sampling produces a function from which observables can be calculated after the simulation. In this section we discuss the challenges associated with using the function $G(n)$ to calculate observables, and the variety of data that we calculate in relation to entanglement.

The first step in using $G(n)$ is normalizing the function, since Wang-Landau sampling converges $G(a) - G(a + 1)$ for all a, but not the absolute magnitude of either. By examining the partition function at infinite temperature we see that

$$Z(\beta = 0) = \sum_{n=0}^{\infty} \beta^n g(n) = g(0), \qquad (14)$$

$$\ln[Z(\beta = 0)] = S(\beta = 0) = G(0), \qquad (15)$$

where $S(\beta = 0)$ is the entropy at infinite temperature. In a simulation employing spin-1/2 particles, at infinite temperature each spin is independent and the entropy is

$$S(\beta = 0) = N_s \ln(2), \qquad (16)$$

with N_s as the number of spins. In the modified simulation with α layers, spins *not* in the region are duplicated $(n - 1)$ times and act as independent degrees of freedom. In this (more general) case, the entropy is

$$S_\alpha(A, \beta = 0) = [N_s + (\alpha - 1)(N_s - N_A)] \ln(2), \qquad (17)$$

where N_A is the number of spins in region A. Using these formulas we can normalize $G(n)$ for any of the simulations by subtracting a constant amount from all elements of $G(n)$ such that $G(0)$ has the correct value.

In addition, if we know the energy at infinite temperature we can gather information about $G(1)$. This is because if we look at the calculation of the energy it can be written as

$$E = -\frac{\sum_n n e^{G(n)} \beta^{n-1}}{\sum_n e^{G(n)} \beta^n} + C, \qquad (18)$$

$$\lim_{\beta \to 0} E = -e^{G(1) - G(0)} + C, \qquad (19)$$

where C is the sum of all constants subtracted from the Hamiltonian in the SSE formalism. This gives us a second check for any model in which the high temperature limit of the energy is known.

Using Eq. (7) and the normalized $G(n)$ we get the equation for the partition function,

$$\ln[Z(\beta)] = \ln\left[\sum_n \beta^n e^{G(n)}\right]. \qquad (20)$$

Although such a calculation does not pose any problems analytically, computing such a large term using fixed precision data types is a small challenge. To calculate it, first we use a second function $G'(\beta, n)$ defined as

$$G'(\beta, n) = G(n) + n \ln(\beta). \qquad (21)$$

Using this function the log of the partition function can be written as

$$\ln[Z(\beta)] = \ln\left[\sum_n e^{G'(\beta, n)}\right]. \qquad (22)$$

Finally, we factor the entire equation by $K = \max_n \{G'(\beta, n)\}$ to get

$$\ln[Z(\beta)] = \ln[K] + \ln\left[\sum_n e^{G'(\beta,n) - K}\right]. \quad (23)$$

By doing this, the series can be reexponentiated, summed, and the natural logarithm taken of the result without the risk of numerical overflow. Other calculations, such as the energy, use a similar trick to prevent the very large values of the partition function from causing problems in computation.

The method so far allows calculation of estimators derived from $G(n)$ but gives no estimate on the *quality* of the data. To determine the method's accuracy (without comparing to known results) we use the fact that, even using the $1/t$ optimization, the initial noise of the simulation is only proven to asymptotically converge. By using multiple independent simulations we can estimate any observable $F[G(n)]$ as

$$\langle F[G(n)]\rangle = \frac{1}{N}\sum_{i=1}^{N} F[G_i(n)], \quad (24)$$

$$\Delta\langle F[G(n)]\rangle = \frac{1}{\sqrt{N}}\left(\left[\frac{1}{N}\sum_{i=1}^{N}(F[G_i(n)])^2\right] \right. $$
$$\left. - \left[\frac{1}{N}\sum_{i=1}^{N} F[G_i(n)]\right]^2\right)^{1/2}, \quad (25)$$

where $\langle\cdots\rangle$ represents the statistical average of a quantity over the simulation and $G_i(n)$ represents the function generated from the ith simulation. The standard deviation of the mean of $\langle F[G(n)]\rangle$ is represented by $\Delta\langle F[G(n)]\rangle$ but does not include systematic errors introduced by features such as the cutoff. We denote the estimator as $\langle F[G(n)]\rangle$ to distinguish our statistical estimate for the observable to the analytically exact $F[G(n)]$ if we had precise knowledge of $G(n)$.

The noise in $G(n)$ might make it tempting to estimate $F[G(n)]$ as

$$\langle F[G(n)]\rangle = F[\langle G(n)\rangle], \quad (26)$$

$$\langle G(n)\rangle = \frac{1}{N}\sum_{i=1}^{N} G_i(n). \quad (27)$$

We have tested that calculating the estimate of the observables by Eq. (24) gives very similar results to Eq. (26), but it does not lend itself easily to calculating the confidence of the estimate. Comparison of $G(n)$ from different simulations shows that the deviation from the mean is highly correlated—that is to say if $G_i(n)$ is larger than average that $G_i(n+k)$ is also likely to be larger than average for small k. This error is clarified by examining the distribution discrete first derivative of $G_i(n)$ in Fig. 2. That the first derivative is the quantity that has a random uncorrelated distribution rather than the function itself can be understood by understanding that it is the ratio between adjacent elements, $g(n+1)/g(n)$ or $G(n+1) - G(n)$, that the Wang-Landau simulation converges for all n and through this the algorithm is able to reconstruct $G(n)$.

III. RESULTS FOR MUTUAL INFORMATION

The Wang-Landau technique outlined above is particularly suited for studying finite-temperature properties of the Rényi entanglement entropies. It is important to note that, at finite temperature, the Rényi entropies in general no longer obey the property that $S_\alpha(\rho_A) = S_\alpha(\rho_B)$, since each quantity picks up volume contributions to its scaling from thermal mixing as the temperature is increased from $T = 0$. The analogous quantity that is studied at finite temperatures is therefore the *mutual information* (MI),

$$I_\alpha = S_\alpha(\rho_A) + S_\alpha(\rho_B) - S_\alpha(\rho_{A\cup B}), \quad (28)$$

which is designed to cancel the volume contributions affecting each Rényi entropy at $T > 0$, resulting in a symmetric quantity that reduces to $I_\alpha = 2S_\alpha(\rho_A) = 2S_\alpha(\rho_B)$ at $T = 0$ assuming $S_\alpha(\rho) = 0$ at zero temperature (which is not true if there is ground-state degeneracy). In QMC simulations, the MI can be calculated for a given region A and its complement (B) using three simulations—two if A and B are congruent. In previous work [15] the calculation for the Rényi entropy was done using thermodynamic integration, extending Eq. (17) to finite temperature by integrating energy, similar to how one would calculate thermodynamic entropy in classical systems:

$$S_\alpha(\rho_A) = -S_\alpha(A, \beta = 0) + \alpha S(\beta = 0)$$
$$+ \int_0^\beta \langle E\rangle_{A,\beta'}\, d\beta' - \alpha \int_0^\beta \langle E\rangle_{\beta'}\, d\beta', \quad (29)$$

where $\langle E\rangle_{A,\beta'}$ $(\langle E\rangle_{\beta'})$ is the energy of the modified simulation cell (normal simulation) at inverse temperature β'. Using this integration technique, the MI was studied using QMC simulation in the context of the anisotropic Heisenberg (or XXZ) model [19]. To calculate results using Eq. (29) simulations were run at many temperatures to provide the value of $\langle E\rangle_{A,\beta'}$ over a sufficiently wide range and fine temperature mesh to perform numerical integration. By studying the finite-size scaling, the form of which is model dependent, of the α-Rényi entropies, the authors discovered that any critical behavior manifest at a finite-temperature phase transition T_c is manifest as approximate crossings of the MI (for different lattice sizes) at T_c and αT_c [19]. In this paper, we calculate the MI using Eq. (28), generating the Rényi entropy using Eqs. (5) and (23).

We first examine the behavior of the MI in the 2D square-lattice XXZ model,

$$H = J\sum_{\langle ij\rangle}\left(\Delta S_i^z S_j^z + S_i^x S_j^x + S_i^y S_j^y\right), \quad (30)$$

using our Wang-Landau algorithm of the previous section, and compare its performance to the conventional integration technique [19]. In addition to the 2D anisotropic Heisenberg model of Ref. [19] $(\Delta = 4)$, we examine the 2D Heisenberg model $(\Delta = 1)$. To test an extended Wang-Landau method, we use the 2D transverse field Ising model [33] in the regime where the transverse field is weak and it has a finite temperature phase to an Ising-like ground state. The Hamiltonian for this model is typically written as

$$H = -J\sum_{\langle ij\rangle}\sigma_i^z\sigma_j^z - \Gamma\sum_i \sigma_i^x, \quad (31)$$

where σ^z and σ^x are Pauli spin operators with eigenvalues ± 1, and we study the model at $\Gamma = J = 1$. Finally, we examine the three-dimensional (3D) Heisenberg model on a cubic lattice [34], which also has a finite-temperature transition.
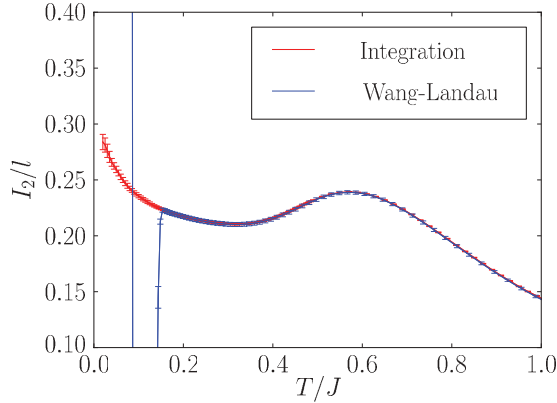
FIG. 3. (Color online) I_2/l, where l is the boundary length, for the 2D Heisenberg model with $L = 12$, comparing the integration of importance sampling and the Wang-Landau algorithm. Notice the sharp anomalous behavior in the Wang-Landau algorithm for temperatures slightly below the cutoff, here chosen to be $T = 0.2$. This means we can only guarantee results down to that temperature, but not that they necessarily diverge there.
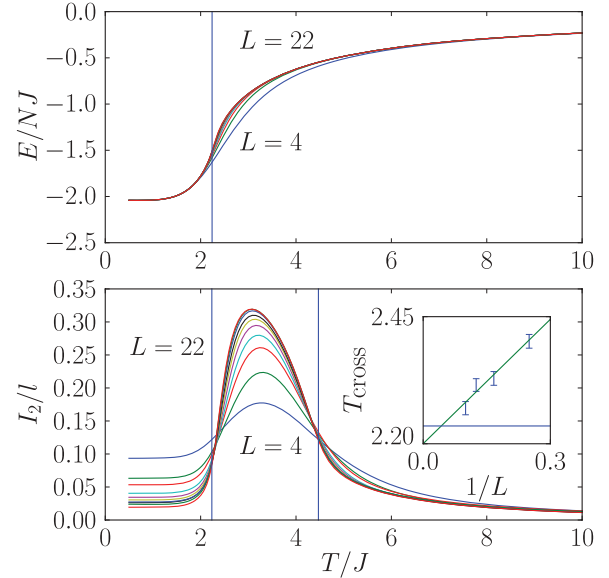

FIG. 4. (Color online) The energy per spin and mutual information divided by boundary length for the 2D XXZ model with $\Delta = 4$. Notice the crossings of the mutual information very close to T_c and $2T_c$, marked with vertical lines. Different lines correspond to lattices of size 4, 6, 8, 10, 12, 14, 16, 18, 20, and 22 (from lowest to highest at peak). The inset shows finite-size extrapolation of crossings of curves from systems of size L and $2L$ (as a function of $1/L$), and gives us $T_c = 2.20(2)$.

To begin, we examine the dependence of the MI on the cutoff error due to our knowledge of $g(n)$ only up to some large (finite) value of n. In the usual SSE QMC simulation the number of operators, n, sampled is fluctuating, but for a given β we can associate an upper bound on the number of operators needed to faithfully represent the system, N. This N can be used to choose an appropriate cutoff in the Wang-Landau algorithm by using a short unmodified SSE QMC simulation to determine the value N. Choosing to have a hard cutoff in n corresponds to a cutoff in reliable data at some β. This can be chosen to be slightly larger than the N identified for the cutoff β in order to ensure that data are still accurate at the β of interest—typically 1.2 to 1.4 times N. Results generated for temperatures above the cutoff are accurate, but those below the cutoff exhibit anomalous behavior, as effectively $g(n) = 0$ for $n > N$ [or equivalently, $G(n) = -\infty$]. The result of attempting to generate results below this cutoff is shown in Fig. 3. $G(n)$ is a smooth function for large enough n, so there is room for methods which generate data using analytic extensions of $G(n)$ to reduce the anomalous behavior near the lower temperature bound, but in general exactly knowing such an analytic form (for all n) would be equivalent to having an analytical expression for the energy as a function of temperature.

We now briefly examine the finite-size scaling of the MI for several examples of Δ in the XXZ model. To begin, recall that the anisotropic Heisenberg model (with $\Delta = 4$) undergoes a phase transition to a z-axis polarized antiferromagnetic ground state as the temperature is lowered. The transition temperature was determined previously using the fourth order binder cumulant of the staggered magnetization, finding $T_c = 2.235(2)$ [19]. The energy per spin and mutual information I_2 divided by boundary size $l = 2L$ is shown in Fig. 4. The mutual information per system size for different L exhibits a universal scaling [19] manifest as approximate crossings at T_c, and $2T_c$ (when using I_2 from the second Rényi entropy). The inset in Fig. 4 shows a closeup of the

crossing, with the uncertainty of the data points calculated from Eq. (25).

It is well known that the 2D isotropic Heisenberg model ($\Delta = 1$) has no finite-temperature phase transition. When examining the mutual information, we do not see any fixed crossing in the curves as a function of L, as shown in Fig. 5; rather, the crossings move towards $T = 0$ for increasing L.

To test the extended broad histogram sampling algorithm [27], we illustrate results from another Hamiltonian, the transverse field Ising model. On the 2D square lattice the classical Ising model has a transition temperature of $T_c = 2J/\ln(1 + \sqrt{2}) \approx 2.269J$ from Onsager's exact result [35]. When we add a small transverse field we expect that the transition temperature will be slightly lowered, and as the critical point of this model for the two-dimensional system is $\Gamma_c = 3.044J$ we are well within the Ising-like phase at $\Gamma = J$ [36]. Figure 6 shows the mutual information curves for this model for a small set of sizes, showing crossings around $T_c \approx 2.21$, slightly below the nonperturbed result, as expected. The inset shows a very tight error bound for the data points when using the extended broad histogram technique.

Finally, we examine the performance of this method in higher dimensions using the 3D isotropic Heisenberg model. Unlike the 2D case, the 3D Heisenberg model does have a finite-temperature phase transition, occurring for $\Delta = 1$ at $T_c = 0.946$ [34]. In this case, the regions A and B used to calculate the Rényi entanglement entropy consisted of a periodic *plane* spanning the 3D simulation cell, cutting it into two halves. MI results are illustrated in Fig. 7 for a largest size of $10 \times 10 \times 10$. In Fig. 7, the MI per surface
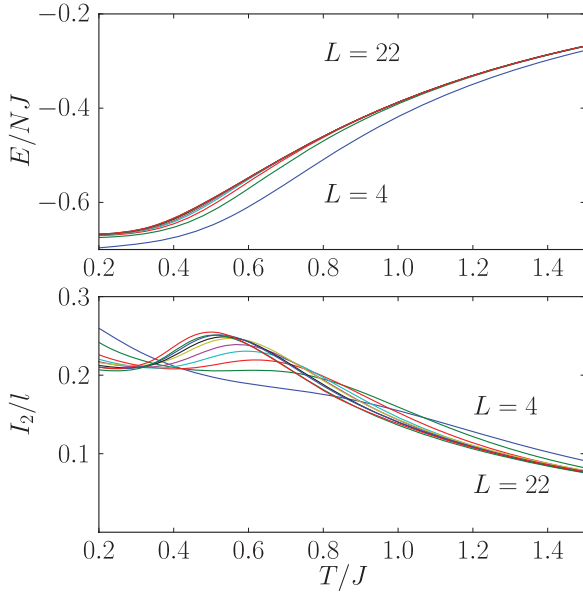
FIG. 5. (Color online) The energy per spin and mutual information divided by system size for the 2D Heisenberg model. This model shows no fixed crossing in I_2/l. Neighboring sizes cross at progressively lower temperatures. Different lines correspond to lattices of size 4, 6, 8, 10, 12, 14, 16, 18, 20, and 22 (from lowest to highest central peak). Crossings cannot be tracked, as for the larger system sizes (i.e., $L = 10$ and $L = 20$) they do not cross at any low temperature.

area of this plane exhibits crossings approaching the critical temperature within error bars. Interestingly, finite-size effects are magnified near $2T_c$, when compared to the 2D transition studied above. It would be interesting to develop a full scaling theory for the 3D Heisenberg transition to examine these effects in the MI, similar to what was done previously in two dimensions [19].
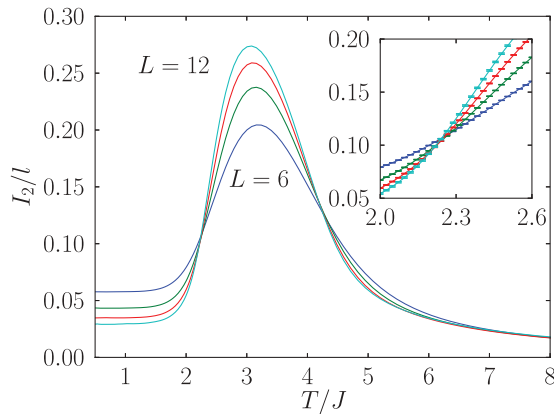

FIG. 6. (Color online) The mutual information divided by system size for the 2D transverse field Ising model using the extended broad sampling approach at $\Gamma = J = 1$. The inset shows a closeup of the lower crossing of mutual information at around $T = 2.21$, crossing which we expect to occur at the transition temperature. Different lines correspond to lattices of size 6, 8, 10, and 12 (from lowest to highest central peak).
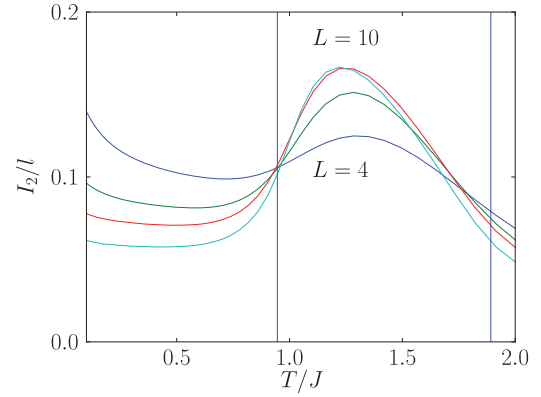

FIG. 7. (Color online) The mutual information divided by surface area (still denoted l) for the 3D Heisenberg model. The crossings of the mutual information occur very close to T_c and to a lesser extent $2T_c$, marked with vertical lines. We believe the deviations would be resolved with a proper finite-size scaling. Different lines correspond to lattices of size 4, 6, 8, and 10 (from lowest to highest central peak).

IV. DISCUSSION

Although Wang-Landau simulations in themselves are not new [21], the use of such noncanonical sampling techniques in quantum Monte Carlo algorithms is still being explored. In general, techniques like Wang-Landau simulations tend to suffer from slower convergence for a specific parameter (i.e., the energy at a specific temperature) when compared to importance sampling techniques, but the trade-off is that the Wang-Landau technique generates this result for all temperatures, something that would require many more simulations using importance sampling. In the end, the decision of the type of sampling to use should depend on the nature of the most important observable.

The Wang-Landau technique can also be extended to measure any quantity that could be measured using traditional SSE. Instead of just simply having $g(n)$, we now associate any number of other observables (like magnetization) with each $g(n)$. By looking at the average value sampled for a given n, when we reweight all the $g(n)$ we can calculate these quantities for any temperature. One difference between Wang-Landau sampling and importance sampling here is that because all energies are explicitly sampled, the simulation does not tend to get stuck in local configurations, although there are certain rare cases in which it can occur. This avoidance of local minimums means that in systems with symmetry broken ground states, like the 2D Ising model, we have a single simulation that is able to access both ground states without the addition of explicit cluster moves.

We have demonstrated that the Wang-Landau technique can be successfully used along with a modified SSE formalism [15] to generate the Rényi entanglement entropies, and associated mutual information, over a wide range of temperatures. Using the modified $1/t$ schedule we are able to reproduce results that agree with previous temperature integration studies on simple quantum models. Improvements using optimized ensembles allow for a similar analysis once the density of states is extracted.

Such broad histogram techniques can also be extended to work in other parameters of the Hamiltonian, such as the size of the entangled region A in the replica trick. A Wang-Landau sampling in this parameter would allow the calculation of the entanglement entropy at a fixed temperature without integration. In fact, the implementation by Humeniuk [20] is suggestive of how one might include Monte Carlo moves that change region size, a necessary element in such a Wang-Landau method. As the accumulation of error is still a concern, it is not trivially clear that integrating in energy or integrating in region size leads to a smaller error for all cases. This should be explored in the future.

Using our Wang-Landau method, we have obtained mutual information results for the 2D anisotropic Heisenberg model, reproducing the crossings at the critical temperature [19]. For the isotropic case, no critical scaling of the mutual information is observed, consistent with the lack of a finite-temperature phase transition in the model. We have also implemented a QMC simulation for the 2D transverse field Ising model, where the mutual information again shows evidence of the expected finite-temperature phase transition. Finally, we examine the mutual information in a three-dimensional Heisenberg model, and show that crossings occur near T_c.

ACKNOWLEDGMENTS

[1] A. Rényi, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, edited by J. Neyman (University of California Press, Berkeley, CA, 1961), pp. 547–561.

[2] C. Holzhey, F. Larsen, and F. Wilczek, Nucl. Phys. B **424**, 443 (1994).

[3] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Phys. Rev. Lett. **90**, 227902 (2003).

[4] E. Fradkin and J. E. Moore, Phys. Rev. Lett. **97**, 050404 (2006).

[5] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006).

[6] M. Levin and X.-G. Wen, Phys. Rev. Lett **96**, 110405 (2006).

[7] M. A. Metlitski, C. A. Fuertes, and S. Sachdev, Phys. Rev. B **80**, 115122 (2009).

[8] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[9] M. B. Hastings, I. González, A. B. Kallin, and R. G. Melko, Phys. Rev. Lett. **104**, 157201 (2010).

[10] Y. Zhang, T. Grover, and A. Vishwanath, Phys. Rev. Lett. **107**, 067202 (2011).

[11] Y. Zhang, T. Grover, and A. Vishwanath, Phys. Rev. B **84**, 075128 (2011).

[12] Y. Zhang, T. Grover, A. Turner, M. Oshikawa, and A. Vishwanath, Phys. Rev. B **85**, 235151 (2012).

[13] J. McMinis and N. M. Tubman, arXiv:1207.4188.

[14] N. M. Tubman and J. McMinis, arXiv:1204.4731.

[15] R. G. Melko, A. B. Kallin, and M. B. Hastings, Phys. Rev. B **82**, 100409 (2010).

[16] P. Calabrese and J. Cardy, J. Stat. Mech.: Theory Exp. (2004) P06002.

[17] Y. Nakagawa, A. Nakamura, S. Motoki, and V. Zakharov, XXVII International Symposium on Lattice Field Theory, Beijing, China, 2009 (unpublished), arXiv:0911.2596.

[18] P. Buividovich and M. Polikarpov, Nucl. Phys. B **802**, 458 (2008).

[19] R. R. P. Singh, M. B. Hastings, A. B. Kallin, and R. G. Melko, Phys. Rev. Lett. **106**, 135701 (2011).

[20] S. Humeniuk and T. Roscilde, Phys. Rev. B **86**, 235116 (2012).

[21] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).

[22] A. W. Sandvik, Phys. Rev. B **59**, R14157 (1999).

[23] O. F. Syljuåsen and A. W. Sandvik, Phys. Rev. E **66**, 046701 (2002).

[24] P. Sengupta, A. W. Sandvik, and D. K. Campbell, Phys. Rev. B **65**, 155113 (2002).

[25] M. Troyer, S. Wessel, and F. Alet, Phys. Rev. Lett. **90**, 120201 (2003).

[26] R. E. Belardinelli, S. Manzi, and V. D. Pereyra, Phys. Rev. E **78**, 067701 (2008).

[27] S. Wessel, N. Stoop, E. Gull, S. Trebst, and M. Troyer, J. Stat. Mech.: Theory Exp. (2007) P12005.

[28] A. W. Sandvik, J. Phys. A **25**, 3667 (1992).

[29] J. E. Hirsch, D. J. Scalapino, R. L. Sugar, and R. Blankenbecler, Phys. Rev. Lett. **47**, 1628 (1981).

[30] N. Kawashima and J. E. Gubernatis, Phys. Rev. Lett. **73**, 1295 (1994).

[31] H. G. Evertz, Adv. Phys. **52**, 1 (2003).

[32] A. W. Sandvik, R. R. P. Singh, and D. K. Campbell, Phys. Rev. B **56**, 14510 (1997).

[33] A. W. Sandvik, Phys. Rev. E **68**, 056701 (2003).

[34] A. W. Sandvik, Phys. Rev. Lett. **80**, 5196 (1998).

[35] L. Onsager, Phys. Rev. **65**, 117 (1944).

[36] H. W. J. Blöte and Y. Deng, Phys. Rev. E **66**, 066110 (2002).