Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations

Matthew B. Hastings,[1] Iván González,[2] Ann B. Kallin,[3] and Roger G. Melko[3]

[1]*Microsoft Research, Station Q, CNSI Building, University of California, Santa Barbara, California, 93106, USA*
[2]*Centro de Supercomputación de Galicia, Avda. de Vigo s/n, E-15705 Santiago de Compostela, Spain*
[3]*Department of Physics and Astronomy, University of Waterloo, Ontario, N2L 3G1, Canada*

(Received 14 January 2010; revised manuscript received 3 March 2010; published 14 April 2010)

We develop a quantum Monte Carlo procedure, in the valence bond basis, to measure the Renyi entanglement entropy of a many-body ground state as the expectation value of a unitary SWAP operator acting on two copies of the system. An improved estimator involving the ratio of SWAP operators for different subregions enables convergence of the entropy in a simulation time polynomial in the system size. We demonstrate convergence of the Renyi entropy to exact results for a Heisenberg chain. Finally, we calculate the scaling of the Renyi entropy in the two-dimensional Heisenberg model and confirm that the Néel ground state obeys the expected area law for systems up to linear size $L = 32$.

DOI: 10.1103/PhysRevLett.104.157201 PACS numbers: 75.10.Jm, 02.70.Ss, 03.65.Ud

The measurement of entanglement entropy provides new tools to determine universal properties of interacting quantum many-body systems in condensed-matter physics. For example, conformally invariant systems in one dimension (1D) display universal logarithmic scaling of entanglement entropy [1]. Many quantum systems in two dimensions (2D) and higher are predicted to have ''area law'' scaling of entanglement entropy, with either universal additive logarithmic corrections for critical systems [2,3] or universal additive constant corrections for topologically ordered systems [4,5].

Scalable simulation methods have become crucial for the study of ground state, finite-temperature, and critical properties of quantum many-body systems. Unfortunately, there are no known scalable simulation methods to calculate entanglement entropy in $D > 1$. Calculating the von Neumann entropy (S_1, defined below) requires calculating the reduced density matrix, which can be done in the Lanczos or density matrix renormalization group (DMRG) schemes. However, scalable simulation techniques in dimensions $D > 1$, are so far essentially restricted to the quantum Monte Carlo (QMC) method, which does not provide access to the reduced density matrix, but only approximates it via a Metropolis importance-sampling algorithm.

We resolve this problem and show that a generalized Renyi entropy (S_2, defined below) can be related to the expectation value of a unitary SWAP operator in the valence bond basis, directly accessible in the QMC method. Using QMC simulations, we calculate S_2 for a spin 1/2 Heisenberg model on 1D and 2D lattices. In 1D, we demonstrate that S_2 calculated with QMC simulations converges to the exact result obtained with DMRG simulations. In 2D, we show that the leading-order term in the scaling of S_2 goes like the boundary of a subregion A, confirming the area law expected from recent studies [6].

The generalized Renyi entanglement entropies are defined by

$$S_n(\rho_A) = \frac{1}{1-n} \ln[\mathrm{Tr}(\rho_A^n)], \tag{1}$$

where ρ_A is the reduced density matrix of a subregion A entangled with its complement B, and $n > 0$. In the limit $n \to 1$, one recovers the familiar von Neumann entropy, $S_1(\rho_A) = -\mathrm{Tr}(\rho_A \ln \rho_A)$. Recently, the generalized Renyi entropies have attracted considerable attention in the condensed-matter community, due to their ability to encode information about the whole ''entanglement spectrum'' of ρ_A, together allowing the set of S_n to contain much more information than S_1 alone [7]. For example, the concept of topological entanglement entropy has recently been generalized to the family of Renyi entropies, where it has been shown to be equal to the logarithm of the total quantum dimension, $S_{\mathrm{top}} \propto \log \mathcal{D}$, independent of n [8]. Universal corrections to the scaling of S_n have also been calculated in a field-theoretic treatment of the $O(N)$ model [9]. Further, any two Renyi entropies S_n and S_m obey the inequality $S_n \geq S_m$ for $n < m$, making S_2 a useful lower bound on S_1.

Valence bond basis and the SWAP *operator.*—We briefly review the notation of Sandvik, and refer the reader to Refs. [10–12] for additional details on the valence bond basis. One begins by writing the expansion of a singlet wave function Ψ_0 of N spins as $N/2$ valence bonds:

$$|\Psi_0\rangle = \sum_r f_r |(a_1^r, b_1^r) \cdots (a_{N/2}^r, b_{N/2}^r)\rangle = \sum_r f_r |V_r\rangle \tag{2}$$

where the valence bonds are singlet pairs denoted by $(a, b) = (|\uparrow_a\downarrow_b\rangle - |\downarrow_a\uparrow_b\rangle)/\sqrt{2}$, occurring between sites a and b of the opposite sublattices of a bipartite lattice, and r labels all possible tilings of bonds. The coefficients f_r are unknown, but an importance-sampling scheme [10] for the valence bond state $|V_r\rangle$ will be outlined below.

To derive an estimator for S_2, we construct two copies of the system (higher entropies S_n require n copies); one "real," one "ancillary" (Fig. 1). Then we define an operator Swap_A which acts between the copies of the system, swapping the configurations within region A. To define this operator, we give its matrix elements in a product basis, e.g., the S^z basis. Let $|\alpha\rangle$ be a complete basis of states in the region A and let $|\beta\rangle$ be a complete basis of states in the complement region B. The state of each copy can be decomposed in this product basis as $|\Psi\rangle = \sum_{\alpha,\beta} C_{\alpha,\beta}|\alpha\rangle|\beta\rangle$ for some amplitudes $C_{\alpha,\beta}$. We define the unitary operator Swap_A by

$$\mathrm{Swap}_A\Big(\sum_{\alpha_1,\beta_1} C_{\alpha_1,\beta_1}|\alpha_1\rangle|\beta_1\rangle\Big)\otimes\Big(\sum_{\alpha_2,\beta_2} D_{\alpha_2,\beta_2}|\alpha_2\rangle|\beta_2\rangle\Big) = \sum_{\alpha_1,\beta_1} C_{\alpha_1,\beta_1}\sum_{\alpha_2,\beta_2} D_{\alpha_2,\beta_2}(|\alpha_2\rangle|\beta_1\rangle)\otimes(|\alpha_1\rangle|\beta_2\rangle). \quad (3)$$

The two copies of the system will not interact with each other, so the ground state of the combined system will be the product of the ground states on each copy: $|\Psi_0\otimes\Psi_0\rangle$. Thus, the expectation value of Swap_A is

$$\langle\Psi_0\otimes\Psi_0|\mathrm{Swap}_A|\Psi_0\otimes\Psi_0\rangle = \sum_{\alpha_1,\alpha_2,\beta_1,\beta_2} C_{\alpha_1,\beta_1}\bar{C}_{\alpha_2,\beta_1}C_{\alpha_2,\beta_2}\bar{C}_{\alpha_1,\beta_2} = \sum_{\alpha_1,\alpha_2}(\rho_A)_{\alpha_1,\alpha_2}(\rho_A)_{\alpha_2,\alpha_1} = \mathrm{Tr}(\rho_A^2), \quad (4)$$

where $(\rho_A)_{\alpha_1,\alpha_2} = \sum_{\beta_1} C_{\alpha_1,\beta_1}\bar{C}_{\alpha_2,\beta_1}$ denotes a matrix element of ρ_A. Therefore,

$$S_2(\rho_A) = -\ln(\mathrm{Tr}(\rho_A^2)) = -\ln(\langle\mathrm{Swap}_A\rangle). \quad (5)$$

Equation (5) is basis independent [13]. In particular it holds in the valence bond basis, where the Swap_A operator is not defined by Eq. (3). Rather, it is defined to swap the endpoints of valence bonds between the real and ancillary copies in the region A, as illustrated in Fig. 1.

Measuring the SWAP *operator in QMC simulations.*—We now present a procedure for measuring the entropy S_2 via importance sampling of the Swap_A operator in QMC simulations. A projector scheme in the valence bond basis has recently been pioneered by Sandvik, and we only briefly review the notation here, referring the reader to Refs. [10–12] for details of implementation. The method is a $T = 0$ projector QMC algorithm, where high powers of a Hamiltonian H are used to project out the ground state $|\Psi_0\rangle$ from a trial wave function Ψ: $(-H)^m|\Psi\rangle \propto |\Psi_0\rangle$. In Sandvik's QMC scheme, $(-H)^m$ is written as a sum of all products of m bond operators,

$$(-H)^m = \sum_r \prod_{i=1}^{m} H_{a_i^r b_i^r} = \sum_r P_r, \quad (6)$$

where for concreteness we use the spin 1/2 Heisenberg model, defining $H = -\sum_{\langle a,b\rangle} H_{ab}$ and $H_{ab} = -(\mathbf{S}_a\cdot\mathbf{S}_b - 1/4)$. The "operator string" P_r is sampled according to its weight, W_r, accrued upon evolution of a trial valence bond state $|V\rangle$ under projection:

$$P_r|V\rangle = W_r|V(r)\rangle. \quad (7)$$

As shown in Ref. [10], for the Heisenberg model, W_r is simply related to the number of off-diagonal operations m_{off} in the projection, $W_r = 2^{-m_{\mathrm{off}}}$.

To sample the Swap_A operator, one requires a double-projector valence bond QMC scheme [10], where a general expectation value for any observable $\mathcal{O}$ is given by

$$\langle\mathcal{O}\rangle = \frac{\sum_{rl}\langle V|P_l^*\mathcal{O}P_r|V\rangle}{\sum_{rl}\langle V|P_l^*P_r|V\rangle} = \frac{\sum_{rl} W_lW_r\langle V(l)|\mathcal{O}|V(r)\rangle}{\sum_{rl} W_lW_r\langle V(l)|V(r)\rangle}. \quad (8)$$

In this case, the two operator strings, P_l and P_r, are applied to two copies of the system; the expectation value of the Swap_A operator as illustrated in Fig. 1 can then be calculated directly. Specifically, one performs importance sampling of operator strings according to the weight $W_lW_r\langle V(l)|V(r)\rangle$, and measures the QMC average expectation value

$$\langle\mathrm{Swap}_A\rangle = \left\langle\frac{\langle V(l)|\mathrm{Swap}_A|V(r)\rangle}{\langle V(l)|V(r)\rangle}\right\rangle, \quad (9)$$

calculating S_2 from Eq. (5).

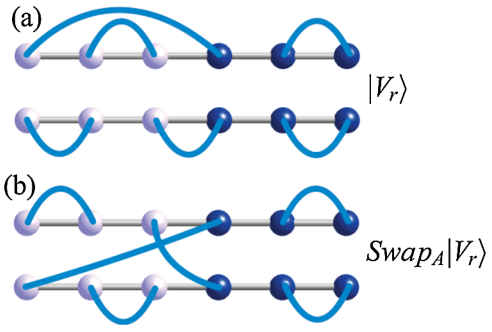

FIG. 1 (color online). A six-site chain, with two noninteracting copies (top and bottom) before (a) and after (b) the Swap_A operation. The region A consists of three light-colored sites on the left; the complement region B of three dark sites on the right. The curved lines denote singlets in the state $|V_r\rangle$, which is a product of two different valence bond states, one per copy. The ground state of the entire system is a linear combination of similar $|V_r\rangle$.

Results for a 1D chain are illustrated in Fig. 2. There, simulations were performed using a double-projector QMC algorithm, with a simple "columnar" trial state $|V\rangle$ (alternating nearest-neighbor bonds in 1D). In the following data, four random bond operators were changed in each operator string per Monte Carlo step, and (unless otherwise stated) $m = 40N$. Exact results for S_2 were obtained with a DMRG simulation, converged with 1000 states. One immediately sees that the naive expectation value $\langle\mathrm{Swap}_A\rangle$ results in very large statistical error bars when the region

A grows large (in 1D, A is the linear region of size i). One may understand this by considering the expected scaling of S_2. Namely, S_2 scales logarithmically with the size of A, so the expectation value $\langle \mathrm{Swap}_A \rangle$ should be polynomially small in A. As a result, there are only very few configurations contributing significantly to $\langle \mathrm{Swap}_A \rangle$, so in practice the importance sampling is done poorly, resulting in large error bars and possibly jeopardizing simulation ergodicity [14]. The problem is even worse in 2D and higher, where the expectation value $\langle \mathrm{Swap}_A \rangle$ should be exponentially small in A.

To combat this issue, we propose a refinement to the algorithm, which we called "improved ratio" sampling. Consider first a one-dimensional chain, and define regions $A^1, A^2, \ldots, A^n$, such that A^i contains i sites, and A^i is a subset of A^{i+1}. In other words, each region A^{i+1} is obtained by adding one site to region A^i, and region A^0 is the empty set (thus, Swap_{A^0} is equal to the identity operator). In a given simulation, one can calculate the ratio

$$\frac{\langle \mathrm{Swap}_{A^{i+1}} \rangle}{\langle \mathrm{Swap}_{A^i} \rangle} = \frac{\sum_{rl} W_l W_r \langle V(l)|\mathrm{Swap}_{A^{i+1}}|V(r)\rangle}{\sum_{rl} W_l W_r \langle V(l)|\mathrm{Swap}_{A^i}|V(r)\rangle} \tag{10}$$

where, for each $i = 0, \ldots, n-1$, the log of this ratio is equal to minus the difference $S_2(\rho_{A^{i+1}}) - S_2(\rho_{A^i})$. Computing this ratio for each i will let us compute S_2 for all A^i. To calculate this ratio, the simulation must be performed with the modified sampling weight,

$$W_l W_r \langle V(l)|\mathrm{Swap}_{A^i}|V(r)\rangle, \tag{11}$$

i.e., a unique QMC simulation must be done for each desired i. If every site were used for a different simulation weight, the QMC algorithm would obtain an additional multiplicative factor of N in its scaling. In practice, this can be reduced by noting that good statistical control can be retained by calculating $\langle \mathrm{Swap}_{A^{i+j}} \rangle / \langle \mathrm{Swap}_{A^i} \rangle$ for fixed i and a range of $j \in [1, j_{\max}]$ (see also the Discussion). We illustrate this in Fig. 2, where simulations with $j_{\max} = 5$ are sufficient to converge the 100-site chain to the exact results along most of its length.

2D Heisenberg results.—We now extend these concepts to QMC simulations of the spin 1/2 Heisenberg model on $N = L \times L$ lattices. For the simulations using the improved ratio estimator, $\langle \mathrm{Swap}_{A^{\ell+r}} \rangle / \langle \mathrm{Swap}_{A^\ell} \rangle$, we define A^ℓ as a square region of linear size ℓ containing $\ell \times \ell$ sites, so $A^{\ell+r}$ contains $2\ell r + r^2$ more sites than A^ℓ.

Figure 3 illustrates S_2 on two 2D Heisenberg models with $L = 8$ and $L = 16$. In the case of $L = 8$, data for a single simulation calculating $\langle \mathrm{Swap}_A \rangle$ directly is essentially identical to the improved ratio estimators for all ℓ. Here, the direct estimator [Eq. (9)] corresponds to one simulation, while $r_{\max} = 1$ corresponds to 6 different simulations of the improved ratio estimator $\langle \mathrm{Swap}_{A^{\ell+1}} \rangle / \langle \mathrm{Swap}_{A^\ell} \rangle$ and $\ell \in [1, 6]$. In contrast to $L = 8$, for $L = 16$ the direct estimator is significantly different than the improved ratio estimator. We can see the convergence of the data as one successively decreases the range of $r \in [1, r_{\max}]$, so that $r_{\max} = 2$ is identical to $r_{\max} = 1$, which is the smallest possible range for the improved ratio estimator in the 2D geometry illustrated. It is important to note that for large $r_{\max}$ the correct value of S_2 does *not* lie within the error bars of the data. This is an indication that simulation ergodicity may be jeopardized by low sampling of the Swap_A operator.

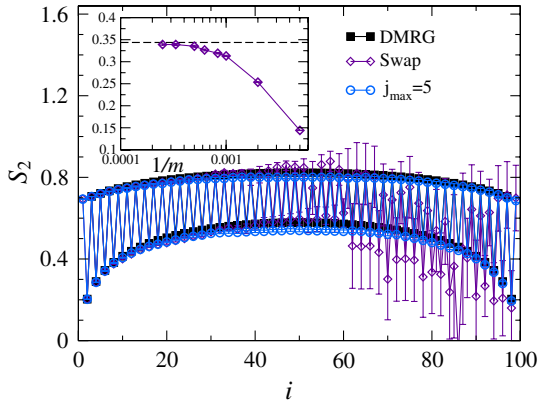

FIG. 2 (color online). The Renyi entropy S_2 as a function of site index $i \in A$, for a 100-site Heisenberg chain with open boundaries, calculated with DMRG and QMC simulations. Data labeled "Swap" was calculated with Eq. (9) with one QMC simulation, while data labeled $j_{\max} = 5$ was calculated with Eq. (10) using 20 separate QMC simulations with a range of $j \in [1, 5]$. The inset shows the convergence of S_2 to the exact value (dashed line) for $i = 6$ with up to $m = 4000$.

We use the improved ratio estimator with $r_{\max} = 1$ to scale the results for the Renyi entropy of the Heisenberg model to larger system sizes in 2D. Results are given in Fig. 4 for $L = 4$ to $L = 32$, plotted as S_2/ℓ. From Ref. [6] one expects the scaling of S_1 in the Néel state to obey the area law; also recall that $S_2 \leq S_1$. The data illustrated is clearly consistent with the area law $S_2/\ell \sim$ const. for $\ell \ll L$, at which point boundary effects likely become important. In particular, multiplicative log corrections (which were apparent in similar system sizes and geometries for the valence bond entanglement entropy [15,16]) are not present.

Discussion.—In this Letter we have presented an algorithm for measuring the Renyi entropy S_2 in valence bond basis QMC simulations via the expectation value of a SWAP operator between two copies of the system. Using an improved ratio estimator, we are able to converge the expectation value of S_2 to the exact result on a 1D Heisenberg chain. Using the same procedure, we have presented the first measurement of the Renyi entanglement entropy in a 2D system, confirming the area law for the Néel ground state of the Heisenberg model.

The simple double-projector QMC algorithm used in this Letter is known to scale approximately as $O(m^2)$,

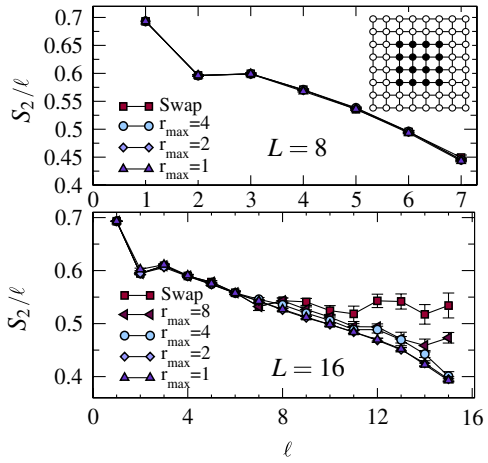

FIG. 3 (color online). The Renyi entropy divided by the linear dimension ℓ of the entangled region A, for 2D lattices with $L = 8$ and $L = 16$. The data labeled "Swap" is from a single simulation calculating Eq. (9) directly. The other data sets are derived from the improved ratio estimator, Eq. (10), with different ranges of $r \in [1, r_{\max}]$ (see text). The inset is a periodic $L = 8$ lattice with region A consisting of the 16 central (dark) sites labeled by $\ell = 4$.

[12] assuming $m > N$. Thus, the current simulation results for the direct SWAP expectation value, Eq. (9), also have $O(m^2)$ scaling, while results obtained using the improved ratio estimator, Eq. (10), have $O(Lm^2)$ scaling in the current geometries. We note however that these geometries may not be ideal for very large ℓ and L, even in the improved ratio sampling. For example, for $L = 32$, results for $r_{\max} = 2$ and $r_{\max} = 1$, although remaining converged within error bars, begin to develop larger discrepancies when ℓ approaches L.

From the current work (which used about 10 CPU years), the question naturally arises whether it is possible to converge the expectation value of the SWAP operator on larger system sizes. We expect that this will be possible

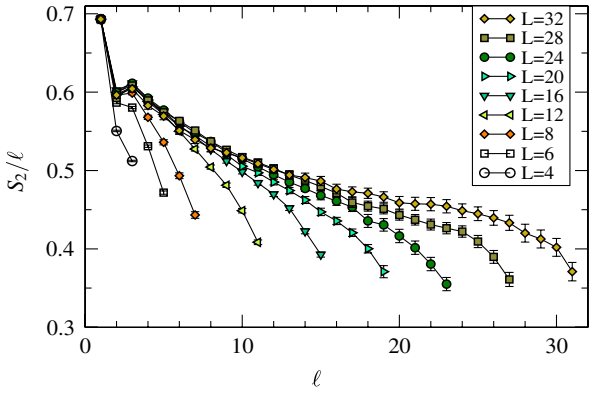

FIG. 4 (color online). Scaling of the Renyi entropy divided by the linear dimension ℓ of the entangled region A for different systems sizes in 2D. All data is calculated with the improved ratio estimator, Eq. (10), with $r_{\max} = 1$.

using a different sequence of regions A^i in the improved ratio estimator. Rather than considering a sequence of regions A^ℓ consisting of $\ell \times \ell$ squares, we can use a more finely grained sequence of regions; e.g., by defining A^i to contain i sites such that for $i = \ell^2$, A^i is an $\ell \times \ell$ square. Incrementing the number of sites in i by one would result in (at worst) $O(Nm^2)$ scaling in the current algorithm. However, we also note that recently-developed *loop* algorithms result in significant improvement in scaling, allowing for the convergence of observables such as energy and spin correlation functions for $L = 256$ and larger [12]. In these schemes, the scaling of the direct SWAP operator [Eq. (9)] would decrease to $O(m)$, while the expected scaling of the improved ratio estimator would be $O(Nm)$ at worst, depending on the definitions of the A^i regions. We therefore expect that the algorithm will require only a polynomial number of samples to converge the Renyi entropy on arbitrary system sizes.

The authors thank A. Sandvik and A. J. Berlinsky for useful discussions. This work was made possible by the computing facilities of SHARCNET and CESGA. Support was provided by NSERC of Canada (A. B. K. and R. G. M.).

[1] P. Calabrese and J. Cardy, J. Stat. Mech. (2004) P06002.
[2] H. Casini and M. Huerta, Nucl. Phys. B **764**, 183 (2007).
[3] S. Ryu and T. Takayanagi, Phys. Rev. Lett. **96**, 181602 (2006).
[4] A. Kitaev and J. Preskill, Phys. Rev. Lett. **96**, 110404 (2006).
[5] M. Levin and X.-G. Wen, Phys. Rev. Lett. **96**, 110405 (2006).
[6] A. B. Kallin, I. González, M. B. Hastings, and R. G. Melko, Phys. Rev. Lett. **103**, 117203 (2009).
[7] H. Li and F. D. M. Haldane, Phys. Rev. Lett. **101**, 010504 (2008).
[8] S. T. Flammia, A. Hamma, T. L. Hughes, and X.-G. Wen, Phys. Rev. Lett. **103**, 261601 (2009).
[9] M. A. Metlitski, C. A. Fuertes, and S. Sachdev, Phys. Rev. B **80**, 115122 (2009).
[10] A. W. Sandvik, Phys. Rev. Lett. **95**, 207203 (2005).
[11] A. W. Sandvik and K. S. D. Beach, *Computer Simulation Studies in Condensed-Matter Physics XX* (Springer, Berlin, 2008), Chap. 22.
[12] A. W. Sandvik and H. G. Evertz, arXiv:0807.0682.
[13] G. Jaeger, *Quantum Information An Overview* (Springer, Berlin, 2007), Chap. 8.
[14] The valence bond basis does alleviate this problem, compared to the S^z basis, in some systems; for example, a valence bond crystal can have an exponentially small $\langle \mathrm{Swap}_A \rangle$, without having *any* realization-to-realization fluctuations, while in the S^z basis the expectation value in a given realization is either 0 or 1.
[15] F. Alet, S. Capponi, N. Laflorencie, and M. Mambrini, Phys. Rev. Lett. **99**, 117204 (2007).
[16] R. W. Chhajlany, P. Tomczak, and A. Wójcik, Phys. Rev. Lett. **99**, 167204 (2007).