Interacting Boson Problems Can Be QMA Hard

Tzu-Chieh Wei,[1,2,3] Michele Mosca,[1,4,5] and Ashwin Nayak[4,1,5]

[1]*Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada*
[2]*Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario, Canada*
[3]*Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia, Canada*
[4]*Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada*
[5]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada*

Computing the ground-state energy of interacting electron problems has recently been shown to be hard for quantum Merlin Arthur (QMA), a quantum analogue of the complexity class NP. Fermionic problems are usually hard, a phenomenon widely attributed to the so-called sign problem. The corresponding bosonic problems are, according to conventional wisdom, tractable. Here, we demonstrate that the complexity of interacting boson problems is also QMA hard. Moreover, the bosonic version of N-representability problem is QMA complete. Consequently, these problems are unlikely to have efficient quantum algorithms.

Many important model Hamiltonians in physics, such as the Hubbard model (both fermionic and bosonic) and those for superconductivity and the quantum Hall effect, involve at most two-body interactions [1]. The ground-state wave function and energy of these Hamiltonians play a key role in understanding these fascinating phenomena. In some of these phenomena, electrons are the major players. Problems involving fermionic particles such as electrons often seem to be computationally more difficult than those with bosonic counterparts. This intractability is often attributed to the so-called "sign problem," occurring in Quantum Monte Carlo simulations [2]. On the other hand, bosonic problems do not suffer from the sign problem [3] and are thus regarded as tractable.

Schuch and Verstraete and Liu, Christandl, and Verstraete have recently shown that computing the ground-state energy of general interacting electrons is QMA hard [4,5]. The complexity class QMA (quantum Merlin Arthur) is a generalization of the class NP (non-deterministic polynomial time) to the quantum realm. It was introduced by Kitaev [6] in the so-called local Hamiltonian problem (LHP), where, roughly speaking, the goal is to determine the ground-state energy of a spin Hamiltonian involving only few-body interaction terms. QMA hard problems are regarded as difficult, unlikely to be solved efficiently even by a quantum computer. However, a quantum computer, if given the solution to any problem in QMA, along with a suitable "certificate" or "witness state," can efficiently verify whether the solution is correct or not. In fact, as a result of a series of works [7–9], even for nearest-neighbor two-body interactions among spin-1/2 particles on a two-dimensional lattice, the LHP is QMA complete. With higher magnitude of spins in one dimension, the LHP can be QMA complete as well [10]. Understanding and classifying the complexity of physical models and investigating hard problems using

statistical mechanical tools have become important research endeavors [4–17], as a result of interplay between physics, chemistry, mathematics, and computer science.

The fact that interacting fermion problems are hard motivates us to investigate the corresponding bosonic problems. Their complexity seems less explored. Could bosonic problems be so hard as to be intractable? We show that generic nearest-neighbor two-body interacting boson problems are indeed QMA hard. Inspired by Ref. [5], we study the bosonic version of the (fermionic) N-representability problem [18,19], where one is given a two-particle reduced density matrix ρ and needs to decide whether there exists a consistent global N-body state σ. This problem has been vastly explored in quantum chemistry [20], as its solution would enable efficient solution of ground-state energy for generic two-body interacting fermionic systems. We show that the bosonic N-representability problem is also QMA complete. Additionally, we show that the bosonic N-representability problem given only diagonal elements is NP hard.

QMA hardness of interacting boson problems.—Consider boson creation and annihilation operators $a_j^\dagger$, a_j for the jth site or mode, obeying the commutations [21]:

$$[a_i, a_j] = 0 = [a_i^\dagger, a_j^\dagger], \qquad [a_i, a_j^\dagger] = \delta_{ij}. \qquad (1)$$

The use of these operators preserves the symmetry of the bosonic wave functions under permutations, and any N-boson wave function can be represented as follows:

$$|\psi\rangle = \sum_{i_1 + \ldots + i_m = N} c_{i_1, \ldots, i_m} (a_1^\dagger)^{i_1} \ldots (a_m^\dagger)^{i_m} |\Omega\rangle, \qquad (2)$$

where i_k ($0 \le i_k \le N$) denotes the number of bosons at the kth site, m is the total number of sites, and $|\Omega\rangle$ denotes the vacuum state without any bosons. Note that we restrict ourselves to states with exactly N bosons [22].

We construct a bosonic Hamiltonian $\mathcal{H}_{\text{bose}}$, whose ground-state energy determines the ground-state energy of the following quantum spin-glass model $\mathcal{H}$:

$$\mathcal{H} = \sum_{\langle i,j \rangle} \sum_{\mu,\nu=0}^{3} c_{ij}^{\mu\nu} \sigma_i^{(\mu)} \otimes \sigma_j^{(\nu)}, \qquad (3)$$

where i, j label the site, $\sigma^{(0)} = \mathbb{1}$ denotes the identity and $\sigma^{(1)} = \sigma^x$, $\sigma^{(2)} = \sigma^y$, and $\sigma^{(3)} = \sigma^z$ are the three Pauli matrices. The coefficients $c_{ij}^{\mu\nu}$ are real but arbitrary. Oliveira and Terhal [9] showed that determining the ground-state energy of $\mathcal{H}$ is QMA hard , even if the interactions are restricted to nearest-neighbor sites $\langle i,j \rangle$ in the two-dimensional square lattice. By way of reduction, solving the ground-state energy of $\mathcal{H}_{\text{bose}}$ is also QMA hard. To construct $\mathcal{H}_{\text{bose}}$, we use the Schwinger boson correspondence between qubit and boson states (see, e.g., Ref. [23]) given by the following map:

$$\sigma_i^x \leftrightarrow a_i^\dagger b_i + b_i^\dagger a_i, \qquad \sigma_i^y \leftrightarrow i(b_i^\dagger a_i - a_i^\dagger b_i),$$
$$\sigma_i^z \leftrightarrow a_i^\dagger a_i - b_i^\dagger b_i,$$

where the operators a_i, b_i correspond to distinct sites. It is easy to verify that the bosonic operators obey the commutation relations of the corresponding Pauli operators. We can regard the qubit at site i as a single boson that can be in one of two different degrees of freedom: $|z_i\rangle \leftrightarrow (a_i^\dagger)^{z_i} \times (b_i^\dagger)^{1-z_i}|\Omega\rangle$ with $z_i \in \{0,1\}$, corresponding to the dual-rail encoding of a photonic qubit in the Knill-Laflamme-Milburn linear-optics quantum computation scheme [24]. Hence, N qubits can be represented by N bosons in $2N$ sites (or N sites with each boson possessing two distinct internal degrees of freedom):

$$|z_1, \ldots, z_N\rangle \leftrightarrow (a_1^\dagger)^{z_1}(b_1^\dagger)^{1-z_1} \cdots (a_N^\dagger)^{z_N}(b_N^\dagger)^{1-z_N}|\Omega\rangle.$$

As any two-local qubit Hamiltonian can be written as a linear combination of terms with at most two Pauli operators, the corresponding bosonic Hamiltonian can be written as a combination of products of at most two annihilation and two creation operators. To ensure that there be exactly one boson on the pair of sites corresponding to i, we add the following extra terms: $P_i \equiv (a_i^\dagger a_i + b_i^\dagger b_i - \mathbb{1})^2$, which commute with other terms in the Hamiltonian. The total bosonic Hamiltonian is then

$$\mathcal{H}_{\text{bose}} \equiv \mathcal{H}(a^\dagger, b^\dagger, a, b) + \sum_i cP_i, \qquad (4)$$

which involves at most nearest-neighbor two-body interactions [25]. By making the weight c large enough, e.g., $\sum_{i,j,\mu,\nu}|c_{ij}^{\mu\nu}|N(N-1)/2$, we guarantee that the ground state of the full Hamiltonian has exactly one boson per site. Thus $\mathcal{H}_{\text{bose}}$ may be represented with at most a polynomially larger number of bits as compared to $\mathcal{H}$. Thus, if one can compute the ground-state energy of general bosonic Hamiltonians with at most two-body interactions, one can solve general spin-1/2 two-local Hamiltonian

problems. As solving the latter is QMA hard, solving the former is QMA hard as well. This shows that interacting boson problems are generally difficult.

QMA hardness of bosonic N-representability problem.— We consider the number of bosons N to be fixed, and assume that the number of modes m that the bosons occupy is large enough, i.e., $m \geq \delta N$ for some constant $\delta > 0$. The number of different ways $\mathcal{N}_m$ that N identical bosons can occupy m sites is $\mathcal{N}_m = \binom{N+m-1}{N}$, which grows exponentially in N, i.e., $\mathcal{N}_m \gtrsim (\delta+1)^N/\delta^N$ when N is large. Given an N-boson state $\rho^{(N)}$, the two-boson reduced state is calculated by tracing out all but two bosons: $\rho^{(2)} \equiv \text{Tr}_{N-2}\rho^{(N)}$, where $\rho^{(N)}$ is in general a mixture of states $|\psi\rangle$ of the form (2). More precisely, $\rho^{(2)}$ is given via its matrix elements:

$$\rho_{ijkl}^{(2)} \equiv \frac{1}{N(N-1)}\langle a_k^\dagger a_l^\dagger a_j a_i \rangle, \qquad (5)$$

where the bracket indicates the expectation value over the state $\rho^{(N)}$. Note that the one-boson reduced density matrix $\rho^{(1)} \equiv \text{Tr}_{N-1}\rho^{(N)}$, defined via $\rho_{ik}^{(1)} \equiv \langle a_k^\dagger a_i \rangle/N$, is completely determined once $\rho^{(2)}$ is known:

$$\rho_{ik}^{(1)} = \sum_l \rho_{ilkl}^{(2)}. \qquad (6)$$

Informally, the bosonic N-representability problem (with m sites) asks whether there is an N-boson state whose two-particle reduced density matrix equals a given state ρ. To deal with technical issues related to precision, we are promised that when there is no N-boson state consistent with it, every two-particle reduced state is "far away" from ρ. Formally, we are given a two-boson density matrix ρ of size $[m(m+1)/2] \times [m(m+1)/2]$, and a real number $\beta \geq 1/\text{poly}(N)$, with all numbers specified with $\text{poly}(N)$ bits of precision. We would like to decide if: ("YES" case) There exists an N-boson state σ such that $\text{Tr}_{N-2}(\sigma) = \rho$, or if ("NO" case). For all N-boson states σ, $\| \text{Tr}_{N-2}(\sigma) - \rho\|_1 \geq \beta$.

We show that the bosonic N representability is QMA hard under Turing reductions [26]. In other words, we show that given an efficient algorithm for bosonic N representability, we can efficiently determine the ground-state energy of $\mathcal{H}_{\text{bose}}$, a QMA hard problem as established above. In the sequel, we refer to an algorithm for N representability as the "membership oracle."

The first step is to write the two-particle interacting terms in $\mathcal{H}_{\text{bose}}$ in terms of a complete orthonormal set $\mathcal{Q}$ of two-particle observables: $H_{\text{two-body}} \equiv \sum_{Q \in \mathcal{Q}} \gamma_Q Q$, where the number of elements $l \equiv |\mathcal{Q}| \sim O(m^4)$. Note that $\text{poly}(m)$ is $\text{poly}(N)$, and so is $\text{poly}(l)$. The observables $\mathcal{Q}$ are constructed as in the fermionic case [5]. We define $a_I \equiv a_{i_2}a_{i_1}$, for all pairs $I = (i_1, i_2)$, $i_1 \leq i_2$. (Note that in the case of fermions $i_1 < i_2$.) We fix a total order (denoted by $\prec$) on pairs of indices I. The observables in $\mathcal{Q}$ are defined as follows:

$$X_{IJ} \equiv \frac{1}{\sqrt{n_I n_J}}(a_I^\dagger a_J + a_J^\dagger a_I), \quad \text{for } I \prec J, \quad (7)$$

$$Y_{IJ} \equiv \frac{-i}{\sqrt{n_I n_J}}(a_I^\dagger a_J - a_J^\dagger a_I), \quad \text{for } I \prec J, \quad (8)$$

$$Z_I \equiv \frac{1}{n_I} a_I^\dagger a_I, \quad \text{for } I \prec L, \quad (9)$$

where the factor $n_I = 1$ if $i_1 \neq i_2$, $n_I = 2$ if $i_1 = i_2$, and L denotes the last pair in the ordering. These operators are Hermitian and the X_{IJ} and Y_{IJ} have expectation values in the interval $[-1, 1]$ and the Z_I have expectation values in $[0, 1]$ for any two-particle state. In the two-particle Hilbert space, they are orthogonal to each other under the trace operator, e.g., $\text{Tr}(X_{IJ}Z_K) = 0$ for all I, J, K. They also form a basis for any two-boson states $\rho^{(2)}$:

$$\rho^{(2)} = Z_L + \sum_{I \prec L} \alpha_{Z_I}(Z_I - Z_L) + \frac{1}{2}\sum_{I \prec J}(\alpha_{(X_{IJ})}X_{IJ}$$
$$+ \alpha_{(Y_{IJ})}Y_{IJ}), \quad (10)$$

where $\alpha_Q = \text{Tr}(Q\rho^{(2)})$ for all $Q \in \mathcal{Q}$. Using Eq. (6) we see that the expectation value $\langle a_i^\dagger a_k \rangle$ of the one-body terms, can be expressed as linear combination of the α_Q. Thus we have $\text{Tr}(H_{\text{bose}}\rho^{(N)}) = \sum_{Q \in \mathcal{Q}} \gamma_Q' \alpha_Q$, where γ_Q' includes the contribution from both one-body and two-body terms, and α_Q are the coefficients of $\rho^{(2)} \equiv \text{Tr}_{N-2}(\rho^{(N)})$. Finding the ground-state energy is equivalent to minimizing the linear function $f(\vec{\alpha}) = \sum_{Q \in \mathcal{Q}} \gamma_Q' \alpha_Q$ subject to the constraint that $\vec{\alpha} \in K_N$, where K_N denotes the convex set of all $\vec{\alpha}$ such that the corresponding state $\rho^{(2)}$ is N representable.

The above minimization of energy, subject to the convex constraint that ρ is N representable, belongs to a class of convex optimization problems which can be solved using the shallow-cut ellipsoid algorithm [27,28] with the aid of an N-representability membership oracle. If K_N is contained in a ball of radius R centered at the origin, and it contains a ball of radius r, the run time is $\text{poly}(\log(R/r))$ and the error in the solution due to computation with finite precision is $1/\text{poly}(R/r)$. The algorithm will be efficient and of polynomially bounded error, if R/r is at most $\text{poly}(l)$.

From the discussion leading to Eq. (10) it follows that K_N is contained in a ball of radius $R = \sqrt{l}$ centered at the origin. However, the method used by Liu *et al.* [5] of regarding a_j as creation operator for a hole in site j cannot be employed to suitably bound r from below. Instead, we explicitly construct a set of N-boson states such that the convex hull of the corresponding vectors $\{\alpha_Q\}$ contains a ball of radius $1/\text{poly}(l)$. Relying on the property that bosons can occupy the same site, for each Q we construct states so as to maximize or minimize α_Q. The resulting points have the property that for any coordinate axis, there exist at least two points at constant distance along that axis. As a consequence, we show that their convex hull (which is

contained in K_N) contains a ball with radius $r \geq 1/\text{poly}(l)$, centered at the center of mass of the points.

An algorithm for bosonic N representability thus enables efficient calculation of the ground-state energy of the Hamiltonian $\mathcal{H}_{\text{bose}}$. Consequently, N representability is QMA hard.

QMA completeness.—Is the bosonic N representability inside QMA or even harder? We show that the bosonic N-representability problem is indeed inside QMA, implying that the problem is QMA complete. To establish this, we construct a QMA proof system, i.e., describe a witness state τ (over polynomially many qubits) for the N representability of a given two-boson density matrix ρ, and a polynomial-time quantum algorithm V (the "verifier") that expects such a pair ρ, τ as input. The verifier determines probabilistically whether a given two-boson density matrix ρ is N representable, or is far from being N representable. In the YES case, V outputs YES with probability p_1, and in the NO case, V outputs NO with probability at least $1 - p_0$. For the problem to be in QMA, the gap $p_1 - p_0$ should be at least $1/\text{poly}(N)$; this gap can be amplified to $1 - e^{-\text{poly}(N)}$ [6,29].

For the witness state, we represent an N-boson state σ using m qudits (d-dimensional quantum systems), via the following correspondence, also known as Holstein-Primakoff bosons (see, e.g., Ref. [23]):

$$a_i \leftrightarrow A_i \equiv \frac{1}{\sqrt{s + S_i^z}} S_i^+, \qquad a_i^\dagger \leftrightarrow A_i^\dagger \equiv \frac{1}{\sqrt{s + S_i^z + 1}} S_i^-,$$

where $S_i^\pm$ are the raising or lowering operators for ith spin, and $2s \geq N$. The spin operators above satisfy the bosonic commutation relations provided the total spin magnitude is s. The boson number states $|n\rangle \in \{|0\rangle, |1\rangle, \ldots, |2s\rangle\}$ at one site correspond to the spin states $|s_n\rangle \in \{|s\rangle, |s - 1\rangle, \ldots, |-s\rangle\}$, and $d = 2s + 1$. A bosonic observable $O = a_i^\dagger a_j^\dagger a_l a_k + a_k^\dagger a_l^\dagger a_j a_i$ is transformed into $\tilde{O} = A_i^\dagger A_j^\dagger A_l A_k + A_k^\dagger A_l^\dagger A_j A_i$, which is a tensor product of at most four single-qudit observables, in contrast to the nonlocal string operators in the fermionic case [5].

The expectation value $\langle \tilde{O} \rangle$ can be estimated efficiently. One method is to explicitly diagonalize the observable $\tilde{O}$ as $\sum_i \lambda_i |\theta_i\rangle\langle\theta_i|$, and measure the given qudit representation $\tilde{\sigma}$ of the bosonic state σ in the basis $\{|\theta_i\rangle\}$. Repeating the measurement on polynomially many copies of $\tilde{\sigma}$, we can estimate $\langle \tilde{O} \rangle$. We note that a quantum circuit using qudits with $d = 2s + 1$ can be implemented efficiently by an equivalent circuit using qubits.

The witness τ consists of polynomially many blocks, where each block has m qudits that represent a state $\tilde{\sigma}$ that is claimed to be an N-boson state with $\text{Tr}_{N-2}(\tilde{\sigma}) = \rho$. The verifier V measures, on each block, the observable $\sum_k A_k^\dagger A_k = ms - \sum_k S_k^z$ to check whether the particle number is N. If not, V outputs NO. This measurement projects each block onto the space of fixed particle number states. If the particle number is N, V continues to perform measurements for a suitable set of observables (e.g., those corre-

sponding to $\mathcal{Q}$) using the projected states. It compares whether the outcomes match the expectation values specified by ρ, to check for consistency. It outputs YES if the errors are less than $\beta/\mathrm{poly}(N)$ (for a suitable polynomial), otherwise outputs NO. When ρ is N representable, the prover supplies polynomially many copies of the correct state σ such that $\mathrm{Tr}_{N-2}(\sigma) = \rho$, and the verifier always answers YES (i.e., $p_1 = 1$), as the measurement outcome is always consistent with ρ. When ρ is not N representable, the prover can cheat by entangling different blocks of qudits. Using a Markov argument, first employed by Aharonov and Regev [30] and later by Liu [28], one can show that the verifier will still output NO with probability $\geq \beta/\mathrm{poly}(N)$. Thus N representability is in QMA [31], and hence QMA complete. This in turn implies that determining the ground-state energy of interacting bosons with two-body interactions is also QMA complete.

Further results.—We follow the same argument as in Ref. [5] and conclude that *pure-state* bosonic N representability is in QMA(k), as the essential point is to verify the purity of the certificate. Next, consider the bosonic N-representability problem when only the diagonal elements $D_{ij} \equiv \langle a_i^\dagger a_j^\dagger a_j a_i \rangle$ are specified. If one considers the case $m = 2N$ and the mapping by the Schwinger representation, one finds that the solution enables one to solve the ground-state energy of local spin Hamiltonians which only contain σ^z operators. The latter corresponds to a classical spin-glass problem, and is known to be NP hard [17]. Thus the problem of deciding N representability given $\{D_{ij}\}$ is also NP hard.

Concluding remarks.—We have shown that two families of boson problems are QMA complete, implying that in the worst-case scenario they are unlikely to be solved efficiently even by quantum computers. However, this does not preclude the possibility of efficient approximation algorithms. Approaches such as mean-field theory, path-integral quantum Monte Carlo calculations [3], and more recently matrix product states [32] and multiscale entanglement renormalization ansatz [33] are important endeavors towards classical approximation algorithms. It is possible that many physical models fall into "easy" instances that can be solved efficiently by these schemes. A more speculative direction is to develop *quantum* approximation algorithms, which have potential speedup.

[1] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders College, Fort Worth, 1976); G. D. Mahan, *Many-Particle Physics* (Kuwer Academic, New York, 2000).

[2] M. Troyer and U.-J. Wiese, Phys. Rev. Lett. **94**, 170201 (2005).

[3] D. M. Ceperley, Rev. Mod. Phys. **67**, 279 (1995).

[4] N. Schuch and F. Verstraete, Nature Phys. **5**, 732 (2009).

[5] Y.-K. Liu, M. Christandl, and F. Verstraete, Phys. Rev. Lett. **98**, 110503 (2007).

[6] A. Yu. Kitaev, A. H. Shen and M. N. Vyalyi, *Classical and Quantum Computation* (AMS, Providence, 2002).

[7] J. Kempe and O. Regev, Quantum Inf. Comput. **3**, 258 (2003).

[8] J. Kempe, A. Kitaev, and O. Regev, SIAM J. Comput. **35**, 1070 (2006).

[9] R. Oliveira and B. Terhal, Quantum Inf. Comput. **8**, 0900 (2008).

[10] D. Aharonov *et al.*, Commun. Math. Phys. **287**, 41 (2009).

[11] A. Kay, Phys. Rev. A **76**, 030307(R) (2007).

[12] S. Bravyi, arXiv:quant-ph/0602108.

[13] D. Nagaj and S. Mozes, J. Math. Phys. (N.Y.) **48**, 072104 (2007).

[14] S. Bravyi *et al.*, Quantum Inf. Comput. **8**, 361 (2008).

[15] S. N. Coppersmith, J. Stat. Phys. **130**, 1063 (2008).

[16] C. R. Laumann *et al.*, Quantum Inf. Comput. **10**, 1 (2010).

[17] F. Barahona, J. Phys. A **15**, 3241 (1982).

[18] A. J. Coleman, Rev. Mod. Phys. **35**, 668 (1963).

[19] C. A. Coulson, Rev. Mod. Phys. **32**, 170 (1960); R. H. Trethold, Phys. Rev. **105**, 1421 (1957).

[20] A. J. Coleman and V. I. Yukalov, *Reduced Density Matrices: Coulson's Challenge* (Springer, Berlin, 2000); J. Cioslowski, *Many-Electron Densities and Reduced Density Matrices* (Springer, Berlin, 2000).

[21] See, e.g., J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Westview Press, Boulder, CO, 1998).

[22] For the purpose of proving QMA completeness, it is technically simpler to consider a fixed number of particles, as the operator norm of the Hamiltonian is then bounded.

[23] A. Auerbach, *Interacting Electrons and Quantum Magnetism* (Springer-Verlag, New York, 1994).

[24] E. Knill, R. Laflamme, and G. Milburn, Nature (London) **409**, 46 (2001).

[25] Terms such as $(a_i^\dagger a_i - b_i^\dagger b_i)$ can occur, e.g., due to the difference in the coupling to the external field that two different hyperfine or spin states (a and b) experience.

[26] C. H. Papadimitriou, *Computational Complexity* (Addison Wesley, Reading, MA, 1994).

[27] M. Grötschel, L. Lovász, and A. Schrijver, *Geometric Algorithms and Combinatorial Optimization* (Springer-Verlag, Berlin, 1988).

[28] Y.-K. Liu, Ph.D. thesis, Unversity of California–San Diego, 2007, arXiv:0712.3041.

[29] C. Marriott and J. Watrous, Comput. Complex. **14**, 122 (2005); D. Nagaj, P. Wocjan, and Y. Zhang, Quantum Inf. Comput. **9**, 1053 (2009).

[30] D. Aharonov and O. Regev, in *Proceedings of FOCS, 2003* (IEEE Press, Los Alamitos, CA, 2003), p. 210, arXiv: quant-ph/0307220. The Markov argument employs randomization, which renders the entanglement across different blocks irrelevant, and one still detects an inconsistency in "NO" case with sufficient probability.

[31] For the discussions of QMA vs QMA$_1$, see S. Aaronson, Quantum Inf. Comput. **9**, 81 (2006).

[32] See, e.g., F. Verstraete, J. I. Cirac, and V. Murg, Adv. Phys. **57**, 143 (2008).

[33] G. Vidal, Phys. Rev. Lett. **101**, 110501 (2008).