Quantum proofs can be verified using only single-qubit measurements

Tomoyuki Morimae,[1,*] Daniel Nagaj,[2,†] and Norbert Schuch[3]

[1]*ASRLD Unit, Gunma University, 1-5-1 Tenjin-cho Kiryu-shi Gunma-ken, 376-0052, Japan*
[2]*Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 84511 Bratislava, Slovakia*
[3]*JARA Institute for Quantum Information, RWTH Aachen University, 52056 Aachen, Germany*

Quantum Merlin Arthur (QMA) is the class of problems which, though potentially hard to solve, have a quantum solution that can be verified efficiently using a quantum computer. It thus forms a natural quantum version of the classical complexity class NP (and its probabilistic variant MA, Merlin-Arthur games), where the verifier has only classical computational resources. In this paper, we study what happens when we restrict the quantum resources of the verifier to the bare minimum: individual measurements on single qubits received as they come, one by one. We find that despite this grave restriction, it is still possible to soundly verify any problem in QMA for the verifier with the minimum quantum resources possible, without using any quantum memory or multiqubit operations. We provide two independent proofs of this fact, based on measurement-based quantum computation and the local Hamiltonian problem. The former construction also applies to QMA_1, i.e., QMA with one-sided error.

I. INTRODUCTION

One of the key questions in computational complexity is to determine the resources required to find a solution to a certain problem. On the other hand, even if we do not know how to produce a solution with some given resources, we can still ask whether those resources allow us to verify the correctness of a given solution. Most importantly, this gives rise to the complexity class NP, the class of decision problems whose "yes" instances have proofs that can be efficiently (in polynomial time, with perfect soundness) verified by a deterministic classical computer. This concept can be generalized by allowing probabilistic verification of proofs, leading to the class Merlin Arthur (MA).

The natural quantum version of these classes is formed by the class quantum Merlin Arthur (QMA) [1–4], which consists of all problems whose "yes" instances have a "quantum proof" (i.e., a quantum state) which can be efficiently verified in polynomial time by a quantum computer. The prototypical QMA-complete problem is to determine the ground-state energy of a Hamiltonian with few-body interactions, known as the local Hamiltonian problem [5] (for a review of recent progress on QMA-complete problems, see [6]). Here, the proof is the ground state itself, whose energy can be efficiently estimated using a quantum computer.

All these classes can be understood in terms of a game between an all-powerful prover Merlin and a rational verifier Arthur with limited resources, where Merlin tries to prove some statement to Arthur by sending him a classical or quantum proof, which Arthur then verifies using a classical (NP, MA) or quantum (QMA) computer. Adding quantum mechanics opens new doors for Merlin to cheat, but on the other hand, performing a quantum computation as a verification procedure gives more power to Arthur as well.

Indeed, since Arthur can always start by measuring in the computational basis, effectively treating the proof as classical, QMA is at least as powerful as NP and MA.

However, will we tip this balance when we restrict Arthur's power and supply him only with restricted quantum resources rather than access to a full quantum computer? Most extremely, we could imagine that Arthur does not have access to a quantum memory and can only perform single-qubit measurements on a sequence of qubits sent by Merlin one by one, possibly in a restricted basis. It seems likely that the class of problems that could be soundly verified in such a setup is much smaller than QMA.

In this paper, we prove that this is not the case: Even if Arthur is limited to single-qubit measurements which are performed one by one on qubits sent sequentially by Merlin, the class of problems which can be proven this way still equals QMA. We will show this in two distinct ways: The first proof utilizes measurement-based quantum computing (MBQC) [7], while the second proof is based on a scheme for single-qubit measurement verification of the ground state for the local Hamiltonian problem [2].

Our first proof uses MBQC, a universal model of quantum computing where adaptive single-qubit measurements are applied to a certain highly entangled many-qubit state, such as the graph state [7]. In our protocol, an honest Merlin sends Arthur the graph state coupled with a witness state, qubit by qubit. As Arthur receives each qubit, he measures it. Using such adaptive single-qubit measurements, he applies the desired QMA verification circuit via a MBQC scheme. On the other hand, if Merlin is malicious, he does not necessarily send the correct graph state. However, Arthur can verify whether he received the proper graph state by measuring a list of stabilizers, which can again be done using only single-qubit measurements, maintaining soundness of the protocol. Our proof relies on the idea of Ref. [8], where graph-state verification is the basis of a multiprover quantum interactive proof system for BQP (bounded quantum polytime) with a classical verifier and several entangled but noncommunicating

*morimae@gunma-u.ac.jp
†daniel.nagaj@savba.sk

provers. Our usage of graph-state verification is much simpler because Arthur does the measurements by himself, and therefore he does not need to test device independence, while his measurement choices are naturally hidden from Merlin. Interestingly, this MBQC approach also allows the reduction of the required *classical* computational ability of Arthur to only XOR (exclusive OR) operations, since it is known [9] that single-qubit unitary operations on a known single-qubit state, single-qubit measurements, and the classical XOR gate are enough for classical universal computing.

In the second proof, instead of implementing a quantum verification circuit, we look at the possibility of determining the ground-state energy of a local Hamiltonian using only single-qubit measurements. The trick is to ask Merlin for the ground state, but to keep secret which of the terms in the Hamiltonian we will measure. In fact, we show how to randomly choose and perform a sequence of single-qubit measurements so that the protocol remains complete and sound. Our single-qubit measurement approach to verifying ground states of a local Hamiltonian was first presented by one of the authors in a Stack-exchange post [10] in response to a question by Lior Eldar. Note that another protocol for this task can be derived from the recent multiprover verification scheme by Ji [11].

The efficiency of our local-Hamiltonian-based protocol is similar (up to polynomial factors) to that of the MBQC approach, which is natural for quantum circuits. However, the MBQC argument also applies to the verification procedure of the complexity class QMA_1, i.e., QMA with perfect completeness, while the Hamiltonian approach does not work there, because it is inherently probabilistic even for an honest Merlin. On the other hand, the local-Hamiltonian approach requires only Pauli measurements, while the MBQC approach involves measurements outside the Clifford basis.

Note that recently another direction of restricting Arthur's power—to Clifford gate operations only—has been considered [12]. It was shown there that even if Arthur's power is restricted to Clifford gate operations, the class equals QMA. The basic tool there is the universality of Clifford gates plus magic states [13]: Merlin sends Arthur many copies of magic states in addition to the witness, and Arthur, who can perform only Clifford gate operations, uses them for universal quantum computing. Even if a malicious Merlin sends some other states pretending to be magic states, Arthur can filter them to guarantee the soundness. The result of this paper, showing that restricting Arthur's power to single-qubit measurements does not weaken the complexity class QMA (or QMA_1), is therefore a step in a similar direction.

Furthermore, Ref. [14] introduced the class stoqMA. It is another variant of QMA, restricting Arthur's ability to classical reversible gates on initial qubits prepared in $|0\rangle$ or $|+\rangle$, and finally measurement of the output qubit in the X basis. Unlike our case, stoqMA is not known to be equal to QMA; we know it is contained in SBP (small bounded-error probability). On the other hand, what if we put restrictions on Merlin instead of Arthur? Reference [15] has shown that even if the quantum witness is restricted to be an equal-weight positive-constants superposition of computational basis states, the class of problems provable in this way is still equal to QMA.

II. PRELIMINARIES

A. QMA and its verification protocol

Consider a language L (i.e., the set of "yes" instances of a problem such as the local Hamiltonian) and denote its instances by x, and the length of the bit string x by $|x|$. The language L belongs to the class QMA(a,b) with $a - b \geqslant 1/\mathrm{poly}(|x|)$ if for each x, there exists a polynomial-size quantum circuit Q_x (from a uniform family of circuits), working on $n = \mathrm{poly}(|x|)$ qubits and $m = \mathrm{poly}(|x|)$ ancilla qubits such that the following is true:

(1) *Completeness*. If $x \in L$, there exists an n-qubit witness state $|\xi_x\rangle$, such that the result of the computational-basis measurement on the first qubit of $Q_x(|\xi_x\rangle \otimes |+\rangle^{\otimes m})$ is 1 with probability $\geqslant a$.

(2) *Soundness*. If $x \notin L$, the result of the computational-basis measurement on the first qubit of $Q_x(|\psi\rangle \otimes |+\rangle^{\otimes m})$ is 1 with probability $\leqslant b$ for any n-qubit input state $|\psi\rangle$.

Usually, each ancilla qubit is initialized in the state $|0\rangle$. Here we choose the basis state $|+\rangle \equiv (|0\rangle + |1\rangle)/\sqrt{2}$ instead, to make the procedure compatible with the standard notation for MBQC.

A priori, the class QMA(a,b) depends on the completeness and soundness parameters a and b. However, there are several ways to amplify [16] the parameter a to make it close to 1 and the parameter b close to 0. In fact, as long as $a - b \geqslant 1/\mathrm{poly}(|x|)$, and $a \leqslant 1 - e^{-\mathrm{poly}(|x|)}, b \geqslant e^{-\mathrm{poly}(|x|)}$, the power of this protocol does not change [4]; we thus simply call QMA(a,b) with these restrictions QMA.

On the other hand, the case $a = 1$ is special, in that it requires the existence of a proof which is accepted with unit probability (perfect completeness); in that case, with the conditions on b as before, the class is denoted by QMA_1.

B. Measurement-based quantum computing

Measurement-based quantum computing (MBQC) is a model of universal quantum computing proposed by Raussendorf and Briegel [7]. In this model, universal quantum computation can be realized by the preparation of a certain many-qubit *resource state*, followed by sequential adaptive single-qubit local measurements on the resource state's qubits. The *cluster state* (or the graph state) [7] is the canonical example of such a resource state. Let $G = (V, E)$ be a graph, where V is the set of vertices and E is the set of edges. The cluster state $|G\rangle$ on the graph G is defined by

$$|G\rangle \equiv \left(\bigotimes_{e \in E} CZ_e \right) |+\rangle^{\otimes |V|},$$

where a qubit in the state $|+\rangle$ is located on each vertex of G, and CZ_e is the controlled-Z gate on the edge e. It is known that, for example, the graph state on the two-dimensional square lattice is a universal resource state [7].

The cluster state can also be used to apply a quantum circuit to an arbitrary input state $|\psi\rangle$. To this end, one splits the vertices into two sets V_1 and V_2. The vertices in V_1 are prepared in $|+\rangle^{\otimes |V_1|}$, while the vertices in V_2 are prepared in $|\psi\rangle$, and subsequently, $\otimes_{e \in E} CZ_e$ is applied along the edges of the graph as shown in Fig. 1 (i.e., a square lattice on V_1 connected to the
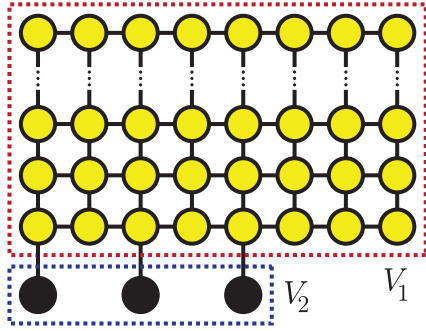
FIG. 1. Graph $G = (V,E)$. Arthur expects Merlin to send him a universal graph state on vertices V_1 and a witness state on vertices V_2, coupled together by controlled-Z gates across the edges between V_1 and V_2.

vertices in V_2). Starting from a state of this form, one can then carry out any desired quantum computation on $|\psi\rangle$ by doing single-qubit measurements only.

III. MBQC APPROACH

Let us now give our first result: a QMA verifier which uses only sequential single-qubit measurements based on MBQC. The basic idea is as follows. Given a QMA problem specified by a verifier circuit Q_x, Arthur asks Merlin to send him the witness state $|\psi\rangle$, coupled to a graph state, which allows Arthur to implement Q_x on $|\psi\rangle$ using MBQC, as in Fig. 1. With an honest Merlin, Arthur can run Q_x on $|\psi\rangle$ using MBQC to verify the proof; on the other hand, we will show that a cheating Merlin can be caught by testing the graph state using only single-qubit measurements. Note that there is no need for Arthur to explicitly ask for ancilla qubits, as these are implicit in a sufficiently large graph state; we can take the ancillas required for computation to be some of the $|+\rangle$ states in the graph state (they are also coupled together with the rest of the graph state by controlled-Z gates).

Let us thus consider a graph $G = (V,E)$ as in Fig. 1. We denote the set of vertices in the red (blue) region by V_1 (V_2), and define $N \equiv |V_1| = \mathrm{poly}(|x|)$. Note that $|V_2| = n$. We also denote the set of edges in the red region by E_1. Let E_{conn} be the set of edges that connect the red and blue regions, i.e., $E_{\mathrm{conn}} = E - E_1$.

Now consider the following interactive proof scheme. Merlin sends a state ρ on G to Arthur, qubit by qubit, in a predefined order. If Merlin is honest, ρ is the graph state plus witness as described above, but if Merlin is malicious, ρ can be any state. Arthur then runs the following protocol. With some probability q, which will be specified later, Arthur uses ρ to run the verifier circuit Q_x on $|\psi\rangle$ using MBQC. If the computation accepts (rejects), Arthur accepts (rejects). On the other hand, with probability $1 - q$, Arthur performs the following *stabilizer test*: He randomly generates an N-bit string $k \equiv (k_1, \ldots, k_N) \in \{0,1\}^N$ and measures the operator

$$s_k \equiv \prod_{j \in V_1} g_j^{k_j},$$

(note that we choose j only from the vertices V_1, and not on the vertices V_2 where the witness is located), where the operator

$$g_j \equiv X_j \bigotimes_{i \in S_j} Z_i$$

is a stabilizer of the graph state which applies X to vertex j, and Z to its neighbors (here, S_j denotes the set of the nearest-neighbor vertices of vertex j). Since s_k is a tensor product of X, Z, and Y, Arthur can (destructively) measure s_k by measuring each qubit independently, as it arrives from Merlin, without the need for memory or multiqubit operations. If the result is $+1$ (-1), the test passes (fails), and Arthur accepts (rejects) Merlin's proof. The probability of passing the stabilizer test is

$$p_{\mathrm{pass}} = \frac{1}{2^N} \sum_{k \in \{0,1\}^N} \mathrm{Tr}\left(\frac{\mathbb{I} + s_k}{2} \rho\right).$$

We will now show that this protocol is complete and sound. First let us consider the case of $x \in L$. Since Merlin is honest, he sends Arthur

$$\left(\bigotimes_{e \in E_{\mathrm{conn}}} CZ_e\right)(|G\rangle_{V_1} \otimes |\xi_x\rangle_{V_2}),$$

where $|G\rangle$ is the graph state on the graph (V_1, E_1), and $|\xi_x\rangle$ is the correct witness state on V_2, which is accepted with probability $\geqslant a$ in the original QMA protocol. The probability of passing the stabilizer test is 1, since we are measuring the stabilizers of a proper graph state. On the other hand, when Arthur chooses to do the computation, he will accept Merlin's proof with probability larger than a (the completeness of the QMA protocol). Therefore, the overall acceptance probability is

$$p_{\mathrm{acc}}^{x \in L} \geqslant qa + (1-q) \equiv \alpha. \qquad (1)$$

Next, let us consider the case $x \notin L$. If Merlin wants to fool Arthur, he has two options: (i) He sends a state that is close to the correct proof (graph state plus witness) and thus has a high probability p_{pass} of passing the stabilizer test; as we will show, such a state will fail the QMA verification. Or (ii) Merlin could try to send a state which is farther away from the correct proof and thus can pass the QMA verification, but such a state will fail the stabilizer test.

Let us thus fix some (small) $\epsilon > 0$ (which will be specified later), and first consider the case where Merlin sends Arthur a state with $p_{\mathrm{pass}} \geqslant 1 - \epsilon$. Then,

$$\mathrm{Tr}\left(\prod_{j \in V_1} \frac{\mathbb{I} + g_j}{2} \rho\right) = \mathrm{Tr}\left(\frac{1}{2^N} \sum_{k \in \{0,1\}^N} s_k \rho\right) \geqslant 1 - 2\epsilon,$$

using the relation

$$\prod_{j \in V_1} \frac{\mathbb{I} + g_j}{2} = \frac{1}{2^N} \sum_{k \in \{0,1\}^N} s_k.$$

Let $W \equiv \bigotimes_{e \in E} CZ_e$ be the product of controlled-Z gates on all edges. Then,

$$
\begin{aligned}
\mathrm{Tr}\left(\prod_{j \in V_1} \frac{\mathbb{I} + g_j}{2} \rho\right) &= \mathrm{Tr}\left(\prod_{j \in V_1} \left(W \frac{\mathbb{I} + g_j}{2} W\right) W \rho W\right) \\
&= \mathrm{Tr}\left(\bigotimes_{j \in V_1} \frac{\mathbb{I} + X_j}{2} W \rho W\right) \\
&= \mathrm{Tr}\left(|+\rangle\langle+|_{V_1}^{\otimes N} \otimes \mathbb{I}_{V_2} W \rho W\right) \\
&= \mathrm{Tr}\left(|+\rangle\langle+|_{V_1}^{\otimes N} \mathrm{Tr}_{V_2}(W \rho W)\right) \\
&= F\left(|+\rangle\langle+|_{V_1}^{\otimes N}, \mathrm{Tr}_{V_2}(W \rho W)\right)^2 \\
&= \max_{w_{V_2}} F\left(|+\rangle\langle+|_{V_1}^{\otimes N} \otimes w_{V_2}, W \rho W\right)^2 \\
&= F\left(W\left(|+\rangle\langle+|_{V_1}^{\otimes N} \otimes w_{V_2}^*\right) W, \rho\right)^2,
\end{aligned}
$$

where Tr_{V_2} is the partial trace over V_2, and $F(\rho, \sigma) \equiv \mathrm{Tr}\sqrt{\sqrt{\rho}\,\sigma\,\sqrt{\rho}}$ is the fidelity between ρ and σ. We have relied on the identity [17]

$$
\begin{aligned}
F\left(|+\rangle\langle+|_{V_1}^{\otimes N}, \mathrm{Tr}_{V_2}(W \rho W)\right) \\
= \max_{w_{V_2}} F\left(|+\rangle\langle+|_{V_1}^{\otimes N} \otimes w_{V_2}, W \rho W\right),
\end{aligned}
$$

and called $w_{V_2}^*$ the state that achieves the maximum. By using the relationship between the trace distance and fidelity, we obtain

$$
\begin{aligned}
\frac{1}{2}\left\| W\left(|+\rangle\langle+|_{V_1}^{\otimes N} \otimes w_{V_2}^*\right) W - \rho \right\|_1 \\
\leqslant \sqrt{1 - F\left(W\left(|+\rangle\langle+|_{V_1}^{\otimes N} \otimes w_{V_2}^*\right) W, \rho\right)^2} \\
\leqslant \sqrt{1 - (1 - 2\epsilon)} = \sqrt{2\epsilon}.
\end{aligned}
$$

Therefore, the acceptance probability for a malicious Merlin that wants to pass the stabilizer test with probability greater than $1 - \epsilon$ is bounded from above by

$$
p_{\mathrm{acc},1}^{x \notin L} \leqslant q(b + \sqrt{2\epsilon}) + (1 - q) \equiv \beta_1. \tag{2}
$$

On the other hand, what happens when Merlin sends Arthur a state that passes the stabilizer test with probability at most $p_{\mathrm{pass}} < 1 - \epsilon$? In that case, we can assume that this false state is far from the graph state, and that Merlin tweaked it in such a way that it passes the computational test with probability = 1. However, the detection probability from the stabilizer test is enough to give us an upper bound on the overall acceptance probability:

$$
p_{\mathrm{acc},2}^{x \notin L} < q + (1 - q)(1 - \epsilon) \equiv \beta_2. \tag{3}
$$

We now need to show that the acceptance probabilities of Eqs. (2) and (3) are necessarily lower than that of Eq. (1), resulting in a completeness-soundness gap for the MBQC-based single-qubit measurement QMA protocol which is lower bounded by an inverse polynomial.

We will do this by finding a setting for q and ϵ that gives us the best completeness-soundness gap. Let us look at the possible gaps:

$$
\Delta_1(q, \epsilon) \equiv \alpha - \beta_1 = qa - q(b + \sqrt{2\epsilon}),
$$

$$
\Delta_2(q, \epsilon) \equiv \alpha - \beta_2 = qa - q + \epsilon(1 - q).
$$

It is optimal for Arthur to choose the value of q that satisfies $\Delta_1 = \Delta_2$:

$$
q^* = \frac{\epsilon}{1 + \epsilon - b - \sqrt{2\epsilon}}.
$$

It is then straightforward to choose $a = \frac{2}{3}$, $b = \frac{1}{3}$ (using amplification for the original circuit), and $\epsilon = \frac{1}{2|x|^2}$, and obtain a completeness-soundness gap for the new MBQC-based protocol

$$
\begin{aligned}
p_{\mathrm{acc}}^{x \in L} - p_{\mathrm{acc}}^{x \notin L} &\geqslant \Delta(q^*, \epsilon) = \frac{\epsilon(a - b - \sqrt{2\epsilon})}{1 + \epsilon - b - \sqrt{2\epsilon}} \\
&\geqslant \frac{\epsilon(a - b - \sqrt{2\epsilon})}{2} = \frac{\frac{1}{3} - \frac{1}{|x|}}{4|x|^2} \geqslant \frac{1}{48|x|^2},
\end{aligned}
$$

for $|x| \geqslant 4$. We have thus proved the new protocol is complete and sound, with an inverse-polynomial promise gap.

Note that this also works for a perfectly complete original QMA protocol with $a = 1$, since the honest acceptance probability remains perfect, $p_{\mathrm{acc}}^{x \in L} = 1$. Therefore, this MBQC-based single-qubit measurement protocol is valid also for QMA_1.

IV. LOCAL-HAMILTONIAN APPROACH

Let us now turn to our second construction for a QMA verification restricted to single-qubit measurements. Rather than implementing the verifier circuit using MBQC, we now devise a way to perform a restricted-quantum-power verification of proofs for a particular QMA-complete problem: the local Hamiltonian [2].

The k-local Hamiltonian promise problem asks whether a Hamiltonian $H = \sum_{m=1}^{M} H_m$ made from k-local (i.e., k-body) terms has a ground-state energy below some E_a, or above some E_b, with a promise gap $E_b - E_a > 1/\mathrm{poly}(|x|)$. (Here, each H_m is bounded.) This problem is QMA complete in short because a successful verification of a proof using a quantum circuit Q_x can be encoded into the ground state of a particular Hamiltonian, and measuring the energy of a state is a simple task using a quantum computer. This can be done either by picking a term H_m at random according to its norm and measuring the expectation value of this term [4] (this procedure can be amplified using multiple copies of the state), or in one go by doing phase estimation of $e^{-iH\tau}$ [18]. Yet both these schemes require joint measurements on at least k qubits. However, as we will show in the following, it is possible to estimate $\langle \psi | H | \psi \rangle$ with single-qubit Pauli measurements only.

The trick is to decompose the Hamiltonian terms in the Pauli basis (or any other local basis), pick one of the Pauli terms according to a particular probability distribution, and measure this term qubit by qubit. Let us present the scheme in full detail. Consider an N-qubit system, and a k-local Hamiltonian $H = \sum_{m=1}^{M} H_m$, together with a promise pair E_a, E_b, separated by $E_b - E_a \geqslant 1/\mathrm{poly}(|x|)$. Each of the terms H_m is k local, acting nontrivially on at most k qubits. We can thus decompose it in the Pauli basis as

$$
H_m = \sum_{S \in \mathcal{P}} c_S^m S,
$$

where $\mathcal{P}$ is the set of tensor products of N Pauli matrices (or identities). We can then rearrange the Hamiltonian H as

$$H = \sum_{S \in \mathcal{P}} \underbrace{\left(\sum_{m=1}^{M} c_S^m \right)}_{d_S} S,$$

labeling d_S the sum of all the prefactors that contribute to the particular k-local Pauli operator S. We now shift the Hamiltonian's spectrum by adding a term proportional to the identity,

$$H' = H + \mathbb{I} \sum_{S \in \mathcal{P}} |d_S|$$
$$= \sum_{S \in \mathcal{P}} |d_S|[\mathbb{I} + \text{sign}(d_S)S] = \sum_{S \in \mathcal{P}} 2|d_S| P_S,$$

which gives us a weighted sum of k-local projectors of the form $P_S = \frac{1}{2}[\mathbb{I} + \text{sign}(d_S)S]$. We now further rescale H', getting

$$H'' = \frac{1}{\sum_S 2|d_S|} H' = \sum_S \pi_S P_S,$$

another weighted sum of projectors, whose weights

$$\pi_S = \frac{|d_S|}{\sum_S |d_S|} \geqslant 0$$

now sum to 1, and thus defines a probability distribution π on a list of at most k-local Pauli operators. The ground state of H is also the ground state of this shifted and rescaled H''. Note that all of these transformations can be carried out classically using only polynomial resources.

Arthur now asks Merlin to send the qubits of the ground state of H one by one. Arthur can estimate the energy of the state $|\psi\rangle$ he receives using the following single-qubit-measurement verification procedure:

(1) Pick S, an at most k-local Pauli product, at random, according to the distribution π.

(2) Ask Merlin to send the qubits of the witness state $|\psi\rangle$ one by one.

(3) On each qubit on which S acts nontrivially, measure the corresponding single-qubit Pauli operator. Take the list of results $x_i = \pm 1$ for $i = 1, \ldots, k$, and calculate the quantity $r = \frac{1}{2}[1 + \text{sign}(d_S)x_1 x_2 \ldots x_k]$, which can take the value 0 or 1.

(4) Accept if $r = 0$.

In the "yes" case, the ground-state energy is promised to be $\leqslant E_a$, and the single-shot probability of obtaining the result 1 when looking at r is

$$p^{\text{yes}}(1) = \langle r \rangle = \langle \psi | H'' | \psi \rangle = \frac{1}{\sum_S 2|d_S|} \langle \psi | H' | \psi \rangle$$
$$= \frac{1}{\sum_S 2|d_S|} \left(\langle \psi | H | \psi \rangle + \sum_S |d_S| \right)$$
$$\leqslant \frac{E_a}{\sum_S 2|d_S|} + \frac{1}{2}, \tag{4}$$

when Merlin sends us a good witness $|\psi\rangle$, whose energy (for the Hamiltonian H) is $\leqslant E_a$. We accept if we measure 0, so

the acceptance probability in this case is at least

$$p_{\text{acc}}^{\text{yes}} = 1 - p^{\text{yes}}(1) \geqslant \frac{1}{2} - \frac{E_a}{\sum_S 2|d_S|}. \tag{5}$$

On the other hand, in the "no" case, all states have energy that is guaranteed to be $\geqslant E_b$. With this in mind, we can just look at Eq. (4) and find a lower bound in terms of E_b. The single-shot probability of measuring the result 1 in our procedure is bounded from below by

$$p^{\text{no}}(1) \geqslant \frac{E_b}{\sum_S 2|d_S|} + \frac{1}{2}.$$

Thus, Arthur will measure $r = 0$ and accept a false proof in this case with probability

$$p_{\text{acc}}^{\text{no}} = 1 - p^{\text{no}}(1) \leqslant \frac{1}{2} - \frac{E_b}{\sum_S 2|d_S|}. \tag{6}$$

Putting together Eqs. (5) and (6) we obtain a gap in the acceptance probabilities between the "yes" and "no" cases:

$$p_{\text{acc}}^{\text{yes}} - p_{\text{acc}}^{\text{no}} \geqslant \frac{E_b - E_a}{\sum_S 2|d_S|}.$$

As the original Hamiltonian H contains $M = \text{poly}(|x|)$ terms H_m of bounded strength, and each H_m has contributions from $4^k = \text{poly}(|x|)$ Pauli products [as long as $k = O(\log |x|)$], we have that $\sum_S |d_S| \leqslant \text{poly}(|x|)$ from above, and thus have obtained a $1/\text{poly}(|x|)$ completeness-soundness gap for our 1-qubit measurement procedure.

This scheme can be used to construct a QMA verification using single-qubit measurements only. Given an instance of a QMA problem, Arthur rewrites the proof as a local Hamiltonian H and asks Merlin for the ground state of H. Using the outlined procedure, Arthur accepts with $p_{\text{acc}}^{\text{yes}}$ and $p_{\text{acc}}^{\text{no}}$, respectively, putting the problem in $\text{QMA}(p_{\text{acc}}^{\text{yes}}, p_{\text{acc}}^{\text{no}}) = \text{QMA}$. Alternatively, this can be seen as a scheme to estimate the ground-state energy with single-qubit measurements only, where multiple copies of the ground state can be used to obtain a more accurate estimate.

Note that the construction presented in this section does not work for QMA_1. For problems in this complexity class one needs to be able to accept perfect witnesses or proofs with 100% probability (perfect completeness). However, our single-qubit-measurement-only procedure for estimating the expectation value of a k-qubit operator is inherently probabilistic for any k-local operator that is not a product of Pauli operators; thus, there is a finite probability that we will reject even a perfect witness.

---

[1] E. Knill, arXiv:quant-ph/9610012.

[2] A. Y. Kitaev, A. H. Shen, and M. N. Vyalyi, *Classical and Quantum Computation* (American Mathematical Society, Boston, 2002).

[3] J. Watrous, in *Proceedings of the 41st Annual Symposium on Foundations of Computer Science* (IEEE Computer Society, Washington, DC, 2000), p. 537.

[4] D. Aharonov and T. Naveh, arXiv:quant-ph/0210077.

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

[6] A. D. Bookatz, Quantum Inf. Comput. **14**, 361 (2014).

[7] R. Raussendorf and H. J. Briegel, Phys. Rev. Lett. **86**, 5188 (2001).

[8] M. McKague, arXiv:1309.5675.

[9] S. Barz, V. Dunjko, F. Schlederer, M. Moore, E. Kashefi, and I. A. Walmsley, arXiv:1501.06730.

[10] N. Schuch, http://cstheory.stackexchange.com/a/12491

[11] Z. Ji, arXiv:1505.07432.

[12] T. Morimae, M. Hayashi, H. Nishimura, and K. Fujii, Quantum Inf. Comput. **15**, 1420 (2015).

[13] S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).

[14] S. Bravyi, A. J. Bessen, and B. M. Terhal, arXiv:quant-ph/0611021.

[15] A. B. Grilo, I. Kerenidis, and J. Sikora, arXiv:1410.2882.

[16] D. Nagaj, P. Wocjan, and Y. Zhang, Quantum Inf. Comput. **9**, 1053 (2011).

[17] M. Hayashi and T. Morimae, Phys. Rev. Lett. **115**, 220502 (2015).

[18] D. Poulin and P. Wocjan, Phys. Rev. Lett. **102**, 130503 (2009).